

(19)



**Евразийское
патентное
ведомство**

(11) **034936**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента
2020.04.08

(21) Номер заявки
201890576

(22) Дата подачи заявки
2016.08.24

(51) Int. Cl. **H04S 1/00** (2006.01)
H04S 3/00 (2006.01)
H04S 7/00 (2006.01)

(54) **КОДИРОВАНИЕ И ДЕКОДИРОВАНИЕ ЗВУКА С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРОВ ПРЕОБРАЗОВАНИЯ ПРЕДСТАВЛЕНИЯ**

(31) **62/209,735; 15189094.4**

(32) **2015.08.25; 2015.10.09**

(33) **US; EP**

(43) **2018.07.31**

(86) **PCT/US2016/048497**

(87) **WO 2017/035281 2017.03.02**

(71)(73) Заявитель и патентовладелец:
**ДОЛБИ ИНТЕРНЕСНЛ АБ
(NL); ДОЛБИ ЛАБОРАТОРИС
ЛАЙСЭНЗИН КОРПОРЕЙШН (US)**

(72) Изобретатель:
**Бребарт Дирк Йерун, Купер Дэвид М.
(AU), Самуэльссон Лейф Дж., Коппенс
Йерун (SE), Уилсон Ронда Джой (US),
Пурнхаген Хейко (SE), Штальман
Александр (DE)**

(74) Представитель:
Носырева Е.Л. (RU)

(56) BREEBAART J. ET AL.: "Multi-channel goes mobile: MPEG surround binaural rendering", AES INTERNATIONAL CONFERENCE. AUDIO FOR MOBILE AND HANDHELDDEVICES, XX, XX, 2 September 2006 (2006-09-02), pages 1-13, XP007902577, page 5 - page 8; figures 8, 11-13
PAULUS JOUNI ET AL.: "MPEG-D Spatial Audio Object Coding for Dialogue Enhancement

(SAOC-DE)", AES CONVENTION 138; MAY 2015, AES, 60 EAST, 42ND STREET, ROOM 2520, NEW YORK, 10165-2520, USA, 6 May 2015 (2015-05-06), pages 10-20, XP040670834, page 11, right-hand column - page 12, left-hand column; figure 1, page 15, left-hand column - page 15, right-hand column; figure 3 page 16, right-hand column

Anonymous: "Dolby AC-4: Audio Delivery for Next-generation Entertainment Services", 1 June 2015 (2015-06-01), XP055324838, Retrieved from the Internet: URL: <http://www.dolby.com/us/en/technologies/ac-4/Next-Generation-Entertainment-Services.pdf> [retrieved on 2016-11-30], page 1 - page 4, page 8 - page 9

"Digital Audio Compression (AC-4) Standard Part 2: Immersive and personalized audio; ETSI TS 103 190-2", ETSI DRAFT; ETSI TS 103 190-2, EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE (ETSI), 650, ROUTE DES LUCIOLES; F-06921, SOPHIA-ANTIPOLIS ; FRANCE, vol. Broadcast, no. V1.1.0, 10 July 2015 (2015-07-10), pages 1-195, XP014249202, [retrieved on 2015-07-10], page 15 - page 24, page 185

US-A1-2014153727

US-A1-2006045294

US-A1-2014355794

J. BREEBAART ET AL.: "Binaural Cues for Multiple Sound Sources", In: "Spatial Audio Processing: MPEG Surround and Other Applications", 1 January 2007 (2007-01-01), John Wiley & Sons, XP055325102, pages 139-154, the whole document

(57) Способ кодирования входного аудиопотока, включающий этапы получения первого представления потока воспроизведения входного аудиопотока, предназначенного для воспроизведения первой системой воспроизведения звука; получения второго представления потока воспроизведения входного аудиопотока, предназначенного для воспроизведения второй системой воспроизведения звука; определения набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, причем параметры преобразования определяются путем минимизации величины разницы между приближением второго представления потока воспроизведения и вторым представлением потока воспроизведения; и кодирования первого представления потока воспроизведения и набора параметров преобразования для передачи в декодер.

B1**034936****034936****B1**

Перекрестная ссылка на родственные заявки

Заявка на данное изобретение заявляет приоритет предварительной заявки на патент США № 62/209735, поданной 25 августа 2015 г., и европейской заявки на патент № 15189094.4, поданной 9 октября 2015 г., которые полностью включаются в данное описание посредством ссылки.

Область техники, к которой относится изобретение

Настоящее изобретение относится к области обработки сигналов, и, в частности, в нем раскрыта система для эффективной передачи звуковых сигналов, содержащих составляющие пространственного представления, т.е. звуковые составляющие, связанные с различными пространственными местоположениями.

Предпосылки изобретения

Любое обсуждение текущего уровня техники в данном описании никоим образом не следует рассматривать как допущение того, что данный уровень широко известен или образует часть общедоступных знаний в данной области.

Создание содержимого, кодирование, распространение и воспроизведение звука обычно выполняется в формате на основе каналов, т.е. одна конкретная целевая система воспроизведения предусмотрена для содержимого во всей экосистеме содержимого. Примерами таких звуковых форматов целевых систем воспроизведения являются монофонический, стереофонический, 5.1, 7.1 и т.д.

Если содержимое подлежит воспроизведению иной системой воспроизведения, нежели предназначенная, может применяться понижающее или повышающее микширование. Например, содержимое в формате 5.1 может быть воспроизведено стереофонической системой воспроизведения путем использования специальных уравнений понижающего микширования. Другим примером является воспроизведение содержимого, кодированного в формате стерео, системой динамиков формата 7.1, которое может включать процесс т.н. повышающего микширования, который может управляться или может не управляться информацией, присутствующей в стереофоническом сигнале. Системой, характеризующейся возможностью осуществления повышающего микширования, является Dolby Pro Logic от Dolby Laboratories Inc (Roger Dressier, "Dolby Pro Logic Surround Decoder, Principles of Operation", www.Dolby.com).

Альтернативной системой звукового формата является формат звуковых объектов, такой как формат, предусмотренный системой Dolby Atmos, см. Robinson, C.Q., Mehta, S., Tsingos, N. (2012), "Scalable format and tools to extend the possibilities of cinema audio" Motion Imaging Journal, SMPTE, 121(8), 63-69. В формате данного типа объектам заданы конкретные местоположения вокруг слушателя, которые могут являться переменными во времени. В таком формате на основе объектов содержимое представляется способом, инвариантным к конкретной системе воспроизведения. Следовательно, для преобразования содержимого в представление, подходящее для конкретной системы воспроизведения, такой как система громкоговорителей или наушники, требуется специализированный процесс представления.

Если стереофоническое, многоканальное или основанное на объектах содержимое подлежит воспроизведению через наушники, часто требуется симитировать многоканальную систему динамиков (для содержимого на основе каналов) или набор виртуальных источников звука (для содержимого на основе объектов) посредством импульсных характеристик для головы (head-related impulse response, HRIR) или бинауральных импульсных характеристик для помещения (binaural room impulse responses, BRIR), имитирующих акустический путь от каждого громкоговорителя к барабанным перепонкам в соответственно безэховой или эховой среде (имитированной). В частности, звуковые сигналы могут быть свернуты с HRIR или BRIR для восстановления интерауральных разностей уровней (inter-aural level difference, ILD), интерауральных временных разностей (inter-aural time difference, ITD) и спектральных меток, позволяющих слушателю определять местоположение каждого отдельного канала. Достижению определенного воспринимаемого расстояния также способствует имитирование акустической среды (реверберация). Обратимся к фиг. 1, на которой показано схематическое общее представление хода обработки данных при представлении двух объектных или канальных сигналов x_i 10, 11, считанных из запоминающего устройства 12 для содержимого для обработки 4 при помощи четырех HRIR, например, 14. Выходные данные HRIR затем суммируются 15, 16 для каждого канального сигнала, чтобы получить выходные данные для наушников для воспроизведения слушателю через наушники 18. Базовый принцип HRIR разъяснен, например, в публикации Wightman, F.L., Kistler, D.J. (1989b). "Headphone simulation of free-field listening. I. Stimulus synthesis", J. Acoust. Soc. Am. 85, 858-867. Результирующий стереофонический сигнал 15, 16 наушников часто называют бинауральным сигналом, бинауральным представлением или (бинауральным) представлением для наушников. Более того, такое бинауральное представление предназначено (или специально рассчитано) для воспроизведения через наушники, в отличие от представления для громкоговорителей, предназначенного для воспроизведения системой громкоговорителей, которая согласуется с каналами, представленными в сигнале (сигналах) представления для громкоговорителей. Эти различные системы воспроизведения называются модальностями, например, одна модальность воспроизведения состоит из наушников, тогда как другая модальность воспроизведения содержит один или более громкоговорителей. Независимо от модальности воспроизведения, на основании входного потока, такого как многоканальный или основанный на объектах формат содержимого, могут быть представлены (сгенерированы) различные представления (стереофоническое, бинауральное, 5.1). В идеа-

ле, для обеспечения того, чтобы художественный замысел был правильно передан слушателю, представления генерируются, или представляются, для конкретных модальностей воспроизведения. В случае воспроизведения через наушники это предполагает применение HRIR или BRIR для создания бинаурального представления, тогда как для громкоговорителей обычно используются методики амплитудного панорамирования. Таким образом, такой процесс представления может применяться как в отношении входного содержимого на основе каналов (5.1, 7.1 и т.п.), так и в отношении погружающего содержимого на основе объектов, такого как Dolby Atmos. В последнем случае в отношении каждого входного объекта обычно независимо используется амплитудное панорамирование (в представлениях для громкоговорителей) или BRIR (в представлениях для наушников) с последующим суммированием вкладов отдельных объектов в результирующий бинауральный сигнал.

Процесс свертки для получения бинаурального представления, предназначенного для воспроизведения наушниками, можно построить так, что метки определения местоположения источника звука, присутствующие в (безэховых) HRTF (head-related transfer functions, передаточных функций головы), восстанавливаются для каждого ввода независимо, в зависимости от (намеченного, воспринимаемого) положения входного канала или объекта, тогда как имитированная эховая среда может быть, по меньшей мере частично, использована совместно по общему алгоритму для двух или большего количества вводов. С этой целью один или более входных сигналов микшируются, или комбинируются, в один или более входных сигналов алгоритма имитирования среды, которые впоследствии обрабатываются для генерирования выходных сигналов имитирования среды, которые могут быть скомбинированы с выходными данными безэхового процесса свертки с использованием HRTF. Алгоритм имитирования среды может имитировать ранние отражения, позднюю реверберацию или и то, и другое, и он может быть реализован посредством известных методик, таких как свертка, задержки, сети задержки обратной связи, всечастотные фильтры и т.п.

Подход на основе свертки с HRIR/BRIR сопровождается несколькими недостатками, одним из которых является существенный объем обработки данных свертки, необходимый для воспроизведения через наушники. Свертку с HRIR или BRIR необходимо применять для каждого входного объекта или канала по отдельности, и, таким образом, сложность обычно возрастает линейно относительно количества каналов или объектов. Поскольку наушники часто используют в сочетании с питающимися от аккумулятора переносными устройствами, высокая вычислительная сложность нежелательна, так как это может существенно сокращать время автономной работы от аккумулятора. Более того, при представлении звукового содержимого на основе объектов, которое может содержать, например, более 100 одновременно активных объектов, сложность свертки с HRIR может быть существенно выше, чем для традиционного содержимого на основе каналов.

Одним из решений для уменьшения вычислительной нагрузки является применение процессов свертки выше по потоку в цепочке обработки данных. Например, в ходе этапа создания содержимого или кодирования. В данном частном случае, называемом "бинауральным предварительным представлением", результирующий бинауральный сигнал, или бинауральное представление, создаваемое в ходе этапа предварительного представления, содержит все метки определения местоположения, предназначенные для воспроизведения наушниками, и в устройстве воспроизведения дополнительная обработка данных не требуется. Недостатком данного способа является то, что вводимые метки определения местоположения источника звука, которые присутствуют в HRIR (такие как интерауральные временные разности (ITD, interaural time differences), интерауральные разности уровней (ILD, interaural level differences), спектральные метки и реверберация), ухудшают воспринимаемое качество при воспроизведении данного конкретного бинаурального представления громкоговорителями, поскольку указанные метки определения местоположения будут применяться фактически дважды: один раз - алгоритмически, на этапе предварительного представления, и один раз - акустически, как следствие акустического пути между громкоговорителями и ушами слушателя.

Краткое описание сущности изобретения

Целью изобретения в его предпочтительном варианте осуществления является создание усовершенствованного варианта осуществления кодирования и декодирования звуковых сигналов для воспроизведения.

В соответствии с первым аспектом настоящего изобретения предусматривается способ кодирования входного аудиопотока, содержащего одну или более звуковых составляющих, причем каждая звуковая составляющая связана с пространственным местоположением, и способ содержит этапы получения первого представления потока воспроизведения для входного аудиопотока, причем первое представление потока воспроизведения представляет собой набор из M1 сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, получения второго представления потока воспроизведения для входного аудиопотока, при этом второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет

собой эховое или безэховое бинауральное представление; определения набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом набор параметров преобразования образует матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, и причем

промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения, при этом параметры преобразования определяются путем минимизации величины разницы между вторым представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представлению потока воспроизведения, и кодирования первого представления потока воспроизведения и набора параметров преобразования для передачи в декодер.

В соответствии со вторым аспектом настоящего изобретения предусматривается способ декодирования представлений потока воспроизведения из потока данных, причем способ включает этапы приема и декодирования первого представления потока воспроизведения, при этом первое представление потока воспроизведения представляет собой набор из $M1$ сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, прием и декодирование набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом набор параметров преобразования образуют матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения; при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление, и при этом промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения, при этом параметры преобразования определяются кодером для обеспечения того, что величина разницы между вторым представленным представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представленному представлению потока воспроизведения, является минимальной, и применения параметров преобразования к промежуточному представлению потока воспроизведения для получения приближения второго представления потока воспроизведения.

В некоторых вариантах осуществления первая система воспроизведения звука может содержать ряд динамиков в фиксированных пространственных местоположениях, а вторая система воспроизведения звука может содержать набор наушников, примыкающих к ушам слушателя. Первое или второе представление потока воспроизведения может представлять собой эховое или безэховое бинауральное представление.

Параметры преобразования предпочтительно являются переменными во времени и зависящими от частоты.

Параметры преобразования предпочтительно определяются путем минимизации величины разницы между результатом применения параметров преобразования к первому представлению потока воспроизведения и вторым представлением потока воспроизведения.

В соответствии с еще одним аспектом настоящего изобретения предусматривается способ кодирования звуковых каналов или звуковых объектов в виде потока данных, включающий этапы приема N входных звуковых каналов или объектов; вычисления набора из M сигналов, где $M \leq N$, путем образования комбинаций N входных звуковых каналов или объектов, при этом набор из M сигналов предназначен для воспроизведения первой системой воспроизведения звука; вычисления набора переменных во времени параметров W преобразования, преобразующих набор из M сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, в приближенное воспроизведение второй системой воспроизведения звука, при этом приближенное воспроизведение приближенно воспроизводит любые эффекты пространственного представления, получаемые путем воспроизведения N входных звуковых каналов или объектов второй системой воспроизведения; и комбинирования M сигналов и параметров W преобразования в поток данных, передаваемый в декодер.

В некоторых вариантах осуществления $M1$ равно 2.

Специалистам в данной области следует принять во внимание, что первый поток представления, кодируемый в кодере, может представлять собой многоканальное представление для громкоговорителей, например окружающее или погружающее (трехмерное) представление для громкоговорителей, такое как

представление в формате 5.1, 7.1, 5.1.2, 5.1.4, 7.1.2 или 7.1.4. В такой ситуации во избежание увеличения вычислительной сложности или для ее минимизации, в соответствии с одним из вариантов осуществления настоящего изобретения, этап определения набора параметров преобразования может включать понижающее микширование первого представления потока воспроизведения в промежуточное представление с меньшим количеством каналов.

В частном примере это промежуточное представление представляет собой двухканальное представление. В данном случае параметры преобразования, таким образом, являются пригодными для преобразования промежуточного двухканального представления во второе представление потока воспроизведения. Первое представление потока воспроизведения может представлять собой окружающее или погружающее представление для громкоговорителей.

Поток данных может дополнительно содержать параметры оценки диалоговых сигналов, при этом способ дополнительно включает: применение параметров оценки диалоговых сигналов к сигналам, предназначенным для воспроизведения первой системой воспроизведения звука, для получения одного или более оценочных диалоговых сигналов; вычитание одного или более оценочных диалоговых сигналов из сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, для получения промежуточного сигнала с подавленным диалогом; применение промежуточного сигнала с подавленным диалогом к процессу имитирования акустической среды для получения одного или более сигналов имитированной акустической среды; и комбинирование одного или более сигналов имитированной акустической среды с аудиопотоком, пригодным для воспроизведения второй системой воспроизведения звука.

Поток данных может дополнительно содержать параметры W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды, при этом способ дополнительно включает: применение параметров W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды к сигналам, предназначенным для воспроизведения первой системой воспроизведения звука, для получения одного или более входных сигналов процесса имитирования акустической среды; применение одного или более входных сигналов процесса имитирования акустической среды к процессу имитирования акустической среды для получения одного или более сигналов имитированной акустической среды; и комбинирование одного или более сигналов имитированной акустической среды с аудиопотоком, пригодным для воспроизведения второй системой воспроизведения звука.

В частном примере первое представленное представление потока воспроизведения представляет собой представление в формате 2.1, и при этом указанный этап понижающего микширования включает преобразование представления в формате 2.1 в стереофоническое представление в формате 2.0 в соответствии с уравнениями:

$$L_o = a*L + b*LFE;$$

$$R_o = a*R + b*LFE,$$

где L , R и LFE обозначают левый и правый каналы полной полосы пропускания и канал низкочастотных эффектов декодированного представления в формате 2.1, при этом a и b представляют собой соответствующие постоянные.

Поток данных предпочтительно может содержать дополнительные параметры W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды, при этом способ дополнительно включает этапы: применения дополнительных параметров W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды к сигналам, предназначенным для воспроизведения первой системой воспроизведения звука, для получения одного или более дополнительных входных сигналов процесса имитирования акустической среды; применения одного или более дополнительных входных сигналов процесса имитирования акустической среды к дополнительному процессу имитирования акустической среды для получения одного или более дополнительных сигналов имитированной акустической среды; и комбинирования одного или более дополнительных сигналов имитированной акустической среды с одним или более из одного или более сигналов имитированной акустической среды и аудиопотоком, пригодным для воспроизведения второй системой воспроизведения звука.

В соответствии с еще одним аспектом настоящего изобретения предусматривается кодер для кодирования входного аудиопотока, содержащего одну или более звуковых составляющих, в котором каждая звуковая составляющая связана с пространственным местоположением, причем кодер содержит

первый блок представления для представления первого представления потока воспроизведения входного аудиопотока, причем первое представление потока воспроизведения представляет собой набор из $M1$ сигналов, предназначенных для воспроизведения первой системой воспроизведения звука;

второй блок представления для представления второго представления потока воспроизведения входного аудиопотока, причем второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление;

блок определения параметров преобразования для определения набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом набор параметров преобразования образует матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, при этом промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения, и при этом параметры преобразования определяются путем минимизации величины разницы между вторым представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представлению потока воспроизведения; и

блок кодирования для кодирования первого представления потока воспроизведения и набора параметров преобразования для передачи в декодер.

В соответствии с еще одним аспектом настоящего изобретения предусмотрен декодер для декодирования представлений потока воспроизведения из потока данных, содержащий блок базового декодера, выполненный с возможностью

приема и декодирования первого представления потока воспроизведения, при этом первое представление потока воспроизведения представляет собой набор из $M1$ сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, и

приема и декодирования набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом набор параметров преобразования образует матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, при этом одно из первого представленного представления потока воспроизведения и второго представленного представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представленного представления потока воспроизведения и второго представленного представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление, и при этом промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения, при этом параметры преобразования определяются кодером для обеспечения того, что величина разницы между вторым представленным представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представленному представлению потока воспроизведения, является минимальной.

Декодер дополнительно содержит матричный множитель для применения параметров преобразования к промежуточному представлению потока воспроизведения для получения приближения второго представления потока воспроизведения.

В соответствии с дополнительным аспектом настоящего изобретения предусмотрен декодер для декодирования ряда звуковых каналов и/или звуковых объектов из потока данных, при этом поток данных содержит набор из M сигналов для воспроизведения первой системой воспроизведения звука и параметры W преобразования, приспособленные для преобразования M сигналов для воспроизведения второй системой воспроизведения звука, причем декодер содержит

блок базового декодера для отделения M сигналов и W параметров преобразования от потока данных, при этом M сигналов разделяется, по меньшей мере, на высоко- и низкочастотные полосы;

матричный множитель для применения W параметров преобразования к M сигналам для получения набора выходных сигналов с частотным разделением; и

блок обратного преобразования, приспособленный для преобразования набора выходных сигналов с частотным разделением в ряд выходных сигналов во временной области, пригодных для воспроизведения второй системой воспроизведения звука.

В некоторых вариантах осуществления декодер может дополнительно содержать блок реверберации, приспособленный для добавления реверберации к набору выходных сигналов с частотным разделением перед преобразованием блоком обратного преобразования.

В некоторых вариантах осуществления первая система воспроизведения звука может содержать набор динамиков, а вторая система воспроизведения звука может содержать набор наушников, при этом параметры W преобразования обеспечивают бинаурализацию набора выходных сигналов с частотным разделением в том смысле, что второе представление потока воспроизведения представляет собой эховое или безэховое бинауральное представление.

В соответствии с дополнительным аспектом настоящего изобретения предусмотрен кодер для кодирования входного аудиопотока, содержащего одну или более звуковых составляющих, при этом каждая звуковая составляющая связана с пространственным местоположением, при этом система содержит

первый блок кодирования для кодирования входного аудиопотока для первой модальности воспроизведения, который выводит первое представление потока воспроизведения; блок определения параметров преобразования для определения ряда параметров преобразования для отображения первого представления потока воспроизведения во второе представление потока воспроизведения; и

второй блок кодирования для кодирования первого представления потока воспроизведения и параметров преобразования в выходной поток кодирования.

Блок определения параметров преобразования определяет ряд параметров преобразования путем минимизации абсолютного значения величины расхождения между требуемым вторым представлением потока воспроизведения и результатом применения ряда параметров преобразования к первому представлению потока воспроизведения.

Термин "ряд" может относиться к свойству наличия переменных во времени параметров преобразования и/или зависящих от частоты параметров преобразования. Второе представление потока воспроизведения может содержать бинаурализованный звук для воспроизведения наушниками.

В соответствии с дополнительным аспектом настоящего изобретения предусмотрен способ получения звукового сигнала для представления наушниками, причем способ включает этапы приема потока данных, содержащего кодированный беззеховый бинауральный сигнал и параметры W (W_F , W_E) генерирования входного сигнала процесса имитирования акустической среды; декодирования кодированного беззехового бинаурального сигнала для получения декодированного беззехового бинаурального сигнала; применения параметров W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды к декодированному беззеховому бинауральному сигналу для получения одного или более входных сигналов процесса имитирования акустической среды; применения одного или более входных сигналов процесса имитирования акустической среды к процессу имитирования акустической среды для получения одного или более сигналов имитированной акустической среды; и комбинирование одного или более сигналов имитированной акустической среды и декодированного беззехового бинаурального сигнала для получения звукового сигнала для представления наушниками.

В некоторых вариантах осуществления один или более сигналов имитированной акустической среды предпочтительно представляют собой один или более из сигналов ранних отражений и сигналов поздней реверберации. Процесс имитирования акустической среды может включать один или более из процесса имитирования ранних отражений и процесса имитирования поздней реверберации. Процесс имитирования ранних отражений может включать обработку одного или более входных сигналов процесса имитирования акустической среды посредством элемента задержки. Процесс имитирования поздней реверберации может включать обработку одного или более входных сигналов процесса имитирования акустической среды посредством сети задержки обратной связи.

Поток данных предпочтительно может содержать дополнительные параметры W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды, при этом способ может дополнительно включать этапы:

применение дополнительных параметров W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды к декодированному беззеховому бинауральному сигналу для получения одного или более дополнительных входных сигналов процесса имитирования акустической среды;

применение одного или более дополнительных входных сигналов процесса имитирования акустической среды к дополнительному процессу имитирования акустической среды для получения одного или более дополнительных сигналов имитированной акустической среды;

комбинирование одного или более дополнительных сигналов имитированной акустической среды с одним или более сигналами имитированной акустической среды и/или декодированным беззеховым бинауральным сигналом.

В соответствии с дополнительным аспектом настоящего изобретения предусмотрен декодер для получения звукового сигнала для представления наушниками, при этом декодер содержит один или более процессоров, выполненных с возможностью

приема потока данных, содержащего кодированный беззеховый бинауральный сигнал и параметры W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды;

декодирования кодированного беззехового бинаурального сигнала для получения декодированного беззехового бинаурального сигнала;

применения параметров W (W_F , W_E) генерирования входных сигналов процесса имитирования акустической среды к декодированному беззеховому бинауральному сигналу для получения одного или более входных сигналов процесса имитирования акустической среды;

применения одного или более входных сигналов процесса имитирования акустической среды к процессу имитирования акустической среды для получения одного или более сигналов имитированной акустической среды;

комбинирования одного или более сигналов имитированной акустической среды и декодированного безэхового бинаурального сигнала для получения звукового сигнала для представления наушниками.

Краткое описание графических материалов

Варианты осуществления изобретения ниже будут описаны с отсылкой к сопроводительным графическим материалам, на которых:

На фиг. 1 показано схематическое общее представление процесса свертки с HRIR для двух объектов источников, в котором каждый канал или объект обрабатывается парой HRIR/BRIR.

На фиг. 2 схематически показано бинауральное предварительно представленное содержимое, воспроизводимое громкоговорителями (известный уровень техники).

На фиг. 3 схематически показано бинауральное предварительно представленное содержимое, воспроизводимое громкоговорителями.

На фиг. 4 схематически показано получение коэффициентов w для обработки представления для громкоговорителей для воспроизведения наушниками.

На фиг. 5 схематически показаны коэффициенты W (W_E), используемые для реконструкции безэхового сигнала и одного раннего отражения (с дополнительным этапом объемной задержки).

На фиг. 6 схематически показан процесс использования коэффициентов W (W_F), используемых для реконструкции безэхового сигнала и входного сигнала FDN из выходных данных базового декодера.

На фиг. 7 схематически показаны получение и обработка коэффициентов w для обработки безэхового представления для наушников и громкоговорителей.

На фиг. 8a, 8b показаны структурные схемы кодера/декодера в соответствии с дополнительным вариантом осуществления настоящего изобретения.

На фиг. 9a показана структурная схема декодера в соответствии с дополнительным вариантом осуществления настоящего изобретения.

На фиг. 9b показана структурная схема упрощенной версии декодера, показанного на фиг. 9a.

Подробное описание

Варианты осуществления предусматривают способ представления звука на основе каналов и/или объектов с низкой битовой скоростью передачи данных и низкой сложностью, который является пригодным для воспроизведения (бинаурального) громкоговорителями и наушниками. Этот способ выполняется путем (1) создания и кодирования представления, предназначенного для воспроизведения конкретной системой воспроизведения (например, без ограничения, громкоговорителями), и (2) добавления дополнительных метаданных, делающих возможным преобразование данного конкретного представления в модифицированное представление, пригодное для другой системы воспроизведения (например, наушников). Данное конкретное представление может называться первым представлением воспроизведения аудиопотока, тогда как модифицированное представление может называться вторым представлением воспроизведения аудиопотока. Первое представление может содержать набор из $M1$ каналов, тогда как второе представление может содержать набор из $M2$ каналов. Количество каналов может являться равным ($M1=M2$) или различным. Метаданные могут иметь форму набора параметров, возможно, переменных во времени и по частоте.

В одной из реализаций метаданные преобразования обеспечивают средства преобразования стереофонического представления для громкоговорителей в бинауральное представление для наушников с возможностью включения ранних отражений и поздней реверберации. Кроме того, для звукового содержимого на основе объектов, виртуальные акустические атрибуты, в частности (относительный) уровень поздней реверберации и/или уровневые, спектральные и временные характеристики одного или более ранних отражений, могут контролироваться для каждого объекта.

Варианты осуществления направлены на исключение артефактов и/или повышение качества воспроизведения и на сохранение художественного замысла при помощи метаданных, управляющих воспроизведением одной или более системами воспроизведения. В частности, варианты осуществления включают метаданные с объектным, канальным или гибридным представлениями сигнала, повышающие качество воспроизведения тогда, когда компоновка системы воспроизведения не соответствует компоновке, предусмотренной в ходе создания содержимого. Как таковое, применение и/или эффект в результате применения метаданных будет зависеть от предназначенных и фактических систем воспроизведения.

Бинауральное предварительно представленное содержимое, воспроизводимое громкоговорителями.

Как описано в разделе о предпосылках, воспроизведение громкоговорителями бинаурального предварительно представленного содержимого может приводить к неестественному тембру из-за того, что спектральные метки, неотъемлемо присутствующие в HRIR или BRIR, применяются дважды: один раз - в ходе предварительного представления, и еще раз - в ходе воспроизведения в акустической среде. Кроме того, каждое воспроизведение бинаурального предварительно представленного содержимого будет неотъемлемо включать азимутальные метки определения местоположения, также применяемые дважды, что вызывает неверное пространственное отображение и ошибки определения местоположения.

На фиг. 2 показан данный вариант осуществления обработки 20. Канал или объект 21 первоначально подвергается свертке 22 с HRIR 23 перед кодированием 25. Таким образом, перед кодированием со-

держимое на основе каналов или объектов подвергается имитированию воспроизведения громкоговорителями посредством обработки HRIR или BRIR. Впоследствии обработанный сигнал кодируется 25, декодируется 26 и воспроизводится громкоговорителями 27, что вносит вышеупомянутые артефакты.

Спектральные артефакты, возникающие в результате двойного применения акустического пути от динамиков к барабанным перепонкам, могут быть, по меньшей мере частично, скомпенсированы путем применения зависящего от частоты коэффициента усиления или ослабления в ходе декодирования или воспроизведения. Данные параметры усиления или ослабления могут впоследствии кодироваться и включаться в содержимое. Для воспроизведения наушниками данные параметры могут быть отброшены, тогда как для воспроизведения громкоговорителями к сигналам перед воспроизведением применяются кодированные коэффициенты усиления.

Один из вариантов осуществления пригодного последовательного хода 30 обработки данных показан на фиг. 3. В данной схеме, когда воспроизведение предназначено для громкоговорителей, метаданные усиления предварительно вычисляются 31 при создании представления. Метаданные являются кодированными с бинаурально обработанными сигналами. В ходе декодирования информации метаданных также декодируется 32. Затем она используется для применения коэффициента 33 усиления к декодированному сигналу для уменьшения значимости артефактов. Для воспроизведения наушниками, с другой стороны, этапы 31-33 не требуются (исключаются), и декодированная информация может применяться непосредственно для воспроизведения наушниками.

Пример реализации.

В одной из реализаций для вычисления метаданных 31 усиления входные сигналы $x_i[n]$ с дискретным временным индексом n и входным индексом i анализируются в частотно-временных мозаиках. Для составления частотно-временных мозаик, каждый из входных сигналов $x_i[n]$ можно разбить на временные кадры, а каждый кадр, в свою очередь, может быть разделен на полосы частот. Полосы частот могут быть получены, например, посредством такого блока фильтров, как блок квадратурных зеркальных фильтров (quadrature mirror filter, QMF), дискретного преобразования Фурье (discrete Fourier transform, DFT), дискретного косинусного преобразования (discrete cosine transform, DCT) или любых других средств разбиения входных сигналов на множество полос частот. Результатом данного преобразования является то, что входной сигнал $x_i[n]$ для входных данных с индексом i и дискретно-временным индексом n представляется сигналами $x_i[k, b]$ подполос для временного интервала (или кадра) k и подполосы b . Краткосрочная энергия в частотно-временной мозаике (K, B) имеет вид:

$$\sigma_{x_i}^2(K, B) = \sum_{b, k \in B, K} x_i[k, b] x_i^*[k, b],$$

где B, K - наборы частотных (b) и временных (k) индексов, соответствующих требуемой частотно-временной мозаике.

Представление бинауральных сигналов $y_l[n]$, $y_r[n]$ в области дискретного времени для левого и правого уха соответственно, имеет вид:

$$y_l[n] = \sum_i x_i[n] * h_{l,i}[n]$$

$$y_r[n] = \sum_i x_i[n] * h_{r,i}[n]$$

где $h_{l,i}[n]$, $h_{r,i}[n]$ - HRIR или BRIR, соответствующая входному индексу i для левого и правого ушей соответственно.

Иными словами, пара $y_l[n]$, $y_r[n]$ бинауральных сигналов может быть создана путем комбинирования свертки и суммирования входных данных i . Впоследствии эти бинауральные сигналы могут быть преобразованы в частотно-временные мозаики с использованием такого же процесса, как процесс, примененный к сигналам $x_i[k, b]$.

Таким образом, для данных бинауральных сигналов в частотной области краткосрочная энергия в частотно-временной мозаике (K, B) может быть вычислена как:

$$\sigma_{y_j}^2(K, B) = \sum_{b, k \in B, K} y_j[k, b] y_j^*[k, b]$$

Далее, на основании сохранения энергии в каждой частотно-временной мозаике, просуммированной по входным объектам i в числителе и по бинауральным сигналам j в знаменателе, могут быть составлены метаданные $w(K, B)$ усиления:

$$w^2(K, B) = \frac{\sum_i \sigma_{x_i}^2(K, B)}{\sum_j \sigma_{y_j}^2(K, B)}$$

Метаданные $w(K, B)$ могут быть затем подвержены квантованию, кодированию и включены в битовый поток аудиокодека. Затем декодер будет применять метаданные $w(K, B)$ к кадру K и полосе B обоих сигналов Y_l и Y_r (входное представление) для получения выходного представления. Такое использование общих метаданных $w(K, B)$, примененных как к Y_l , так и к Y_r , обеспечивает сохранение стереофонического баланса входного представления. Помимо вышеописанного способа, в котором бинауральные сигналы $y_l[n]$, $y_r[n]$ создаются посредством свертки во временной области, процесс бинаурального представления является также применимым в частотной области. Иными словами, вместо первоначального вычисления бинауральных сигналов $y_l[n]$, $y_r[n]$ во временной области, можно вместо этого преобразовать входные сигналы $x_i[n]$ в представление в частотной области и применить процесс свертки с HRIR в частотной области для генерирования представления бинауральных сигналов $y_j[k, b]$ в частотной области, например, при помощи способов быстрой свертки в частотной области. При таком подходе представление бинауральных сигналов $y_j[k, b]$ в частотной области получается без необходимости в генерировании этих сигналов во временной области, и для него не требуется применение блока фильтров или преобразование для применения в отношении бинауральных сигналов во временной области.

Воспроизводимое наушниками стереофоническое содержимое, включая безэховое бинауральное представление.

В данной реализации стереофонический сигнал, предназначенный для воспроизведения громкоговорителями, кодируется в сочетании с дополнительными данными для способствования воспроизведению данного сигнала для громкоговорителей через наушники. Если дан набор входных объектов или каналов $x_i[n]$, то набор сигналов $z_s[n]$ для громкоговорителей обычно генерируется посредством коэффициентов $g_{i,s}$ усиления амплитудного панорамирования, представляющих коэффициент усиления объекта i в динамике s :

$$z_s[n] = \sum_i g_{i,s} x_i[n]$$

Для содержимого на основе каналов коэффициенты $g_{i,s}$ усиления амплитудного панорамирования обычно являются постоянными, тогда как для содержимого на основе объектов, в котором предполагаемое положение объекта представляется при помощи переменных во времени метаданных, коэффициенты усиления, соответственно, будут переменными во времени.

Если даны сигналы $z_s[n]$, подлежащие кодированию и декодированию, то набор коэффициентов w требуется найти так, чтобы, если эти коэффициенты применяются к сигналам $z_s[n]$, то результирующие модифицированные сигналы \hat{y}_l, \hat{y}_r , составленные как:

$$\hat{y}_l = \sum_s w_{s,l} z_s$$

$$\hat{y}_r = \sum_s w_{s,r} z_s$$

точно соответствовали бинауральному представлению первоначальных входных сигналов $x_i[n]$ в соответствии с выражениями:

$$y_l[n] = \sum_i x_i[n] * h_{l,i}[n]$$

$$y_r[n] = \sum_i x_i[n] * h_{r,i}[n]$$

Коэффициенты w можно найти путем минимизации L2-критерия E между требуемым и фактическим бинауральными представлениями:

$$E = |y_l - \hat{y}_l|^2 + |y_r - \hat{y}_r|^2$$

$$w = \arg \min(E)$$

Решение для минимизации ошибки E можно получить при помощи решений в замкнутом виде, методов градиентного спуска или любого другого подходящего итеративного метода минимизации функции ошибок. В качестве одного примера такого решения можно записать различные этапы представле-

ния в матричной записи:

$$Y = XH$$

$$Z = XG$$

$$\hat{Y} = XGW = ZW$$

Данная матричная запись основана на одноканальном кадре, содержащем N дискретных значений, представленных в виде одного столбца:

$$\vec{x}_i = \begin{bmatrix} x_i[0] \\ \vdots \\ x_i[N-1] \end{bmatrix}$$

и матриц как комбинаций нескольких каналов $i = \{1, \dots, I\}$, каждый из которых представлен в матрице одним вектором-столбцом:

$$X = [\vec{x}_1 \quad \dots \quad \vec{x}_I]$$

Решение для W , которое минимизирует E , имеет вид:

$$W = (G^*X^*XG + \epsilon I)^{-1}G^*X^*XN,$$

где $(*)$ - оператор комплексно-сопряженного транспонирования,

I - единичная матрица и

ϵ - постоянная регуляризации.

Данное решение отличается от способа на основе коэффициентов усиления тем, что сигнал \hat{Y} генерируется матрицей, а не скаляром W , применяемым к сигналу Z , что предусматривает возможность наличия скрещивающихся членов (например, второго сигнала \hat{Y} , являющегося (частично) реконструированным из первого сигнала Z).

В идеале коэффициенты w определяются для каждой частотно-временной мозаики для минимизации ошибки E в каждой частотно-временной мозаике.

В приведенных выше фрагментах описания для определения матричных коэффициентов использовался критерий минимальной среднеквадратичной ошибки (L2-критерий). Без потери общности, вместо или в дополнение к принципу минимальной среднеквадратичной ошибки, аналогично могут быть использованы другие хорошо известные критерии или методы вычисления матричных коэффициентов. Например, матричные коэффициенты могут быть вычислены с использованием членов более высокого порядка или путем минимизации L1-критерия (например, критерия наименьшего абсолютного отклонения). Кроме того, могут быть использованы разнообразные методы, включающие методики неотрицательного разложения или оптимизации, непараметрические оценки, оценки максимального правдоподобия и т.п. В дополнение, матричные коэффициенты могут быть вычислены с использованием итеративных процессов или процессов градиентного спуска, методов интерполяции, эвристических методов, динамического программирования, машинного обучения, нечеткой оптимизации, имитированной закалки, или могут быть использованы решения в замкнутом виде и методики анализа через синтез. И последнее, но не менее важное, оценка матричных коэффициентов может быть ограничена различными способами, например, при помощи ограничения диапазона значений, регуляризирующих членов, суперпозиции требований сохранения энергии и т.п.

В ситуациях практического применения HRIR или BRIR $\mathbf{h}_{i,j}, \mathbf{h}_{e,i}$ будет включать зависящие от частоты задержки и/или сдвиги по фазе. Соответственно, коэффициенты w могут являться комплекснозначными с мнимой составляющей, по существу, отличной от нуля.

Один из вариантов реализации обработки данных в данном варианте осуществления показан 40 на фиг. 4. Звуковое содержимое 41 обрабатывается блоком 42 анализирующих гибридных комплексных квадратурных зеркальных фильтров (hybrid complex quadrature mirror filter, HCQMF) в сигналы подполос. Затем, для генерирования бинауральных сигналов Y , к выходным данным блока фильтров применяются 43 HRIR 44. Параллельно входные данные представляются 45 для воспроизведения громкоговорителями, в результате чего образуются сигналы Z громкоговорителей. В дополнение, из сигналов Z громкоговорителей и бинауральных сигналов Y вычисляются 46 коэффициенты (или весовые коэффициенты) w , причем они включаются в битовый поток 48 базового кодера. Могут быть использованы различные базовые кодеры, такие как, например, MPEG-1 Layer 1, 2, 3, раскрытые в публикации Brandenburg, K., & Bosi, M. (1997), "Overview of MPEG audio: Current and future standards for low bit-rate audio coding". Journal of the Audio Engineering Society, 45(1/2), 4-21 or Riedmiller, J., Mehta, S., Tsingos, N., & Boon, P. (2015). "Immersive and Personalized Audio: A Practical System for Enabling Interchange, Distribution, and Delivery of Next-Generation Audio Experiences". Motion Imaging Journal, SMPTE, 124(5), 1-23, которые включаются посред-

ством ссылки. Если базовый кодер не обладает способностью использования сигналов подполос в качестве входных данных, эти сигналы подполос могут сначала быть преобразованы во временную область с использованием блока 47 синтезирующих гибридных комплексных квадратурных зеркальных фильтров (HCQMF).

На стороне декодирования, если декодер выполнен с возможностью воспроизведения для наушников, коэффициенты извлекаются 49 и применяются 50 к сигналам базового декодера перед синтезом 51 HCQMF и воспроизведением 52. Если базовый кодер не вырабатывает сигналы в области HCQMF, может потребоваться необязательный блок 54 анализирующих фильтров HCQMF, как показано на фиг. 4. Таким образом, сигналы, кодированные базовым кодером, предназначены для воспроизведения громкоговорителями, тогда как коэффициенты преобразования сигналов громкоговорителей в бинауральные сигналы определяются в кодере и применяются в декодере. Декодер может быть дополнительно оснащен функциональной возможностью коррекции пользователем и, таким образом, в режиме воспроизведения для наушников пользователь может выбирать воспроизведение наушниками обычных сигналов громкоговорителей, а не бинаурально обработанных сигналов. В данном случае декодер игнорирует весовые коэффициенты. Наконец, если декодер настроен на воспроизведение через громкоговорители, весовые коэффициенты могут быть проигнорированы, и сигналы базового декодера могут воспроизводиться системой воспроизведения через громкоговорители либо непосредственно, либо после повышающего микширования или понижающего микширования для соответствия компоновке системы воспроизведения через громкоговорители.

Будет очевидно, что способы, описанные в предшествующих параграфах, не ограничиваются использованием блоков квадратурных зеркальных фильтров, поскольку могут с не меньшим успехом использоваться как блоки фильтров другой конструкции, так и краткосрочные дискретные преобразования Фурье с обработкой методом окна.

Данная схема обладает различными преимуществами по сравнению с традиционными подходами. Эти преимущества включают следующие:

1) сложность декодера лишь в самой малой степени превышает сложность обычно стереофонического воспроизведения, поскольку дополнение в декодер состоит только из простой (зависящей от времени и частоты) матрицы, управляемой при помощи информации битового потока;

2) данный подход является пригодным для содержимого на основе каналов и на основе объектов, и он не зависит от количества объектов или каналов, присутствующих в содержимом;

3) HRTF становятся параметрами настройки кодера, т.е. они могут быть модифицированы, улучшены, изменены или приспособлены в любой момент времени вне зависимости от совместимости декодера. Для декодеров, представленных в данной области техники, HRTF могут по-прежнему быть оптимизированы или настроены в соответствии с требованиями пользователя без необходимости в модификации этапов обработки данных на стороне декодера;

4) битовая скорость передачи данных является чрезвычайно низкой по сравнению с битовыми скоростями передачи данных, необходимыми для многоканального содержимого или содержимого на основе объектов, поскольку из кодера необходимо передать лишь небольшое количество сигналов громкоговорителей (как правило, один или два) с дополнительными данными (с низкой скоростью передачи) для коэффициентов w ;

5) один и тот же битовый поток может быть верно воспроизведен громкоговорителями и наушниками;

6) битовый поток может быть составлен масштабируемым образом; если, в контексте конкретной услуги, в конечной точке гарантировано использование только громкоговорителей, то коэффициенты w могут быть удалены из битового потока без последствий для традиционного представления громкоговорителями;

7) отличительные признаки передовых кодеков, действующие в отношении представлений для громкоговорителей, такие как управление громкостью, усиление диалога и т.д., будут продолжать действовать по назначению (при воспроизведении громкоговорителями);

8) за счет масштабирования коэффициентов w , громкость бинаурального представления может обрабатываться независимо от громкости воспроизведения громкоговорителями;

9) слушатели, использующие наушники, могут выбирать прослушивание бинаурального или традиционного стереофонического представления вместо принудительного прослушивания того или другого.

Расширение с ранними отражениями.

Часто для повышения реалистичности бинаурального представления, требуется включение в бинауральные сигналы одного или более ранних отражений, используемых вследствие наличия пола, стен или потолка. Если отражение имеет бликовую природу, то оно может быть само по себе интерпретировано как бинауральное представление, в котором соответствующие HRIR содержат эффект поглощения поверхностью, увеличения задержки и пониженный общий уровень из-за увеличения длины акустического пути от источника звука к барабанным перепонкам.

Эти свойства могут быть зарегистрированы модифицированной схемой, такой как схема, показанная на фиг. 5, которая представляет собой модификацию схемы, показанной на фиг. 4. В кодере 64

коэффициенты W определяются для (1) реконструкции безэхового бинаурального представления из представления для громкоговорителей (коэффициенты W_Y), и (2) для реконструкции бинаурального представления отражения из представления для громкоговорителей (коэффициенты W_E). В данном случае безэховое бинауральное представление определяется с помощью HRIR H_e бинаурального представления, что приводит к паре Y безэховых бинауральных сигналов, тогда как раннее отражение определяется с помощью HRIR H_e , приводящими к паре сигналов E раннего отражения. Для того чтобы сделать возможной параметрическую реконструкцию раннего отражения из микшированного стереофонического сигнала, важно, чтобы задержка, вызванная большей длиной пути раннего отражения, была удалена из HRIR H_e в кодере и чтобы данная конкретная задержка была применена в декодере.

Декодер будет генерировать пару безэховых сигналов и пару сигналов раннего отражения путем применения коэффициентов W (W_Y ; W_E) к сигналам громкоговорителей. Для имитирования большей длины пути для раннего отражения, раннее отражение впоследствии обрабатывается посредством этапа 68 задержки. Параметр задержки блока 68 может быть включен в битовый поток кодера или может представлять собой параметр, определяемый пользователем, или он может быть сделан не зависящим от имитированной акустической среды или зависящим от фактической акустической среды, в которой находится слушатель.

Расширение с поздней реверберацией.

Для включения в бинауральное представление имитирования поздней реверберации может быть использован такой алгоритм поздней реверберации, как сеть задержки обратной связи (FDN, feedback-delay network). FDN принимает в качестве входных данных один или более объектов или каналов и вырабатывает (в случае бинаурального ревербератора) два сигнала поздней реверберации. В качестве входных данных в FDN в традиционном алгоритме могут быть использованы выходные данные декодера (или результат его понижающего микширования). Данный подход обладает значительным недостатком. Во многих случаях использования может потребоваться регулировка величины поздней реверберации для каждого объекта. Например, при уменьшении величины поздней реверберации повышается разборчивость диалога.

В одном из альтернативных вариантов осуществления управление величиной реверберации для каждого объекта или каждого канала может создаваться аналогично тому, как из стереофонического микшированного сигнала составляются безэховое бинауральное представление или бинауральное представление ранних отражений.

Как показано на фиг. 6, для дополнительного приспособления к поздней реверберации могут быть выполнены различные модификации предыдущих схем. В кодере 81 вычисляется 82 входной сигнал F FDN, который может представлять собой взвешенную комбинацию входных данных. Указанные весовые коэффициенты могут зависеть от содержимого, например, в результате разметки вручную в ходе создания содержимого или автоматической классификации при помощи интеллектуальных алгоритмов анализа медиаданных. Сам входной сигнал FDN отбрасывается блоком 83 оценки весовых коэффициентов, но данные W_F коэффициентов, делающие возможной оценку, реконструкцию или приближение входного сигнала FDN из представления для громкоговорителей включаются 85 в битовый поток. В декодере 86 входной сигнал FDN реконструируется 88, обрабатывается самой FDN и включается 89 в бинауральный выходной сигнал для слушателя 91.

В дополнение, FDN может быть составлена так, что она допускает несколько (два или более) входов, и, таким образом, пространственные величины входных сигналов сохраняются на выходе FDN. В этих случаях в битовый поток включаются данные коэффициентов, делающие возможной оценку каждого входного сигнала FDN из представления для громкоговорителей.

В этом случае может потребоваться управление определением пространственного местоположения объекта или канала относительно входных данных FDN.

В некоторых случаях может быть возможно генерирование входных сигналов имитирования поздней реверберации (например, FDN) в ответ на параметры, присутствующие в потоке данных для другого назначения (например, параметров, специально не предназначенных для применения к базовым сигналам для генерирования входных сигналов FDN). Например, в одной примерной системе усиления диалога диалоговый сигнал реконструируется из набора базовых сигналов путем применения к этим базовым сигналам параметров усиления диалога. Диалоговый сигнал затем улучшается (например, усиливается) и микшируется обратно в базовые сигналы (что, таким образом, усиливает диалоговые составляющие относительно остальных составляющих базовых сигналов). Как описано выше, часто требуется составление входного сигнала FDN таким образом, чтобы он не содержал диалоговые составляющие. Таким образом, в системах, где уже доступны параметры усиления диалога, можно реконструировать требуемый входной сигнал FDN без диалога (или, по меньшей мере, с подавленным диалогом) путем, в первую очередь, реконструкции диалогового сигнала из базового сигнала и параметров усиления диалога, а затем вычитания (например, аннулирования) диалогового сигнала из базовых сигналов. В такой системе специализированные параметры для реконструкции входного сигнала FDN из базовых сигналов могут не являться обязательными (так как вместо них могут быть использованы параметры усиления диалога) и, таким образом, могут быть исключены, что приводит к уменьшению необходимой скорости передачи

данных параметров без потери функциональных возможностей.

Комбинирование ранних отражений и поздней реверберации.

Несмотря на то что в предшествующих разделах обозначены расширения безэхового представления за счет раннего отражения (отражений) и поздней реверберации, также возможны их комбинации. Например, система может содержать:

1) коэффициенты W_Y для определения безэхового представления из представления для громкоговорителей;

2) дополнительные коэффициенты W_E для определения определенного количества ранних отражений из представления для громкоговорителей;

3) дополнительные коэффициенты W_F для определения одного или более входных сигналов поздней реверберации из представления для громкоговорителей, что позволяет управлять величиной поздней реверберации для каждого объекта.

Безэховое представление как первое представление.

Несмотря на то, что использование представления для громкоговорителей в качестве первого представления, подлежащего кодированию базовым кодером, обладает преимуществом обеспечения обратной совместимости с декодерами, не обладающими способностью интерпретации или обработки данных в преобразования, указанное первое представление не ограничено представлением для воспроизведения громкоговорителями. На фиг. 7 показано схематическое представление способа 100 кодирования и декодирования звукового содержимого 105 для воспроизведения наушниками 130 или громкоговорителями 140. Кодер 101 принимает входное звуковое содержимое 105 и обрабатывает эти сигналы при помощи блока 106 фильтров HCQMF. Впоследствии на основании базы 104 данных HRIR/HRTF элементом 109 свертки с HRIR генерируется безэховое представление Y . В дополнение, элементом 108, вычисляющим и применяющим матрицу G панорамирования громкоговорителей, вырабатывается представление Z для громкоговорителей. Кроме того, элемент 107 вырабатывает входной микшированный сигнал F_{FDN} .

Безэховый сигнал Y , необязательно, преобразовывается во временную область с использованием блока 110 синтезирующих фильтров HCQMF и кодируется базовым кодером 111. Блок 114 оценки преобразования вычисляет параметры W_F (112), делающие возможной реконструкцию входного сигнала F_{FDN} из безэхового представления Y , а также параметры W_Z (113) для реконструкции представления Z для громкоговорителей из безэхового представления Y . Параметры 112 и 113 включаются в битовый поток базового кодера. В качестве альтернативы или в дополнение, несмотря на то, что это не показано на фиг. 7, блок оценки преобразования может вычислять параметры W_E , делающие возможной реконструкцию сигнала E раннего отражения из безэхового представления Y .

Декодер имеет два режима работы, показанных как режим 102 декодера, предназначенный для прослушивания 130 через наушники, и режим 103 декодера, предназначенный для воспроизведения 140 громкоговорителями. В случае воспроизведения наушниками базовый декодер 115 декодирует безэховое представление Y и декодирует параметры W_F преобразования. Впоследствии параметры W_F преобразования применяются к безэховому представлению Y при помощи блока 116 матрицирования для получения оценочного входного сигнала F_{FDN} , который впоследствии обрабатывается F_{FDN} 117 для получения сигнала поздней реверберации. Этот сигнал поздней реверберации микшируется с безэховым представлением Y при помощи сумматора 150, после которого следует блок 118 синтезирующих фильтров HCQMF, для получения представления 130 для наушников. Если также имеются параметры W_E , декодер может применять эти параметры к безэховому представлению Y для получения оценочного сигнала раннего отражения, который впоследствии обрабатывается посредством задержки и микшируется с безэховым представлением Y .

В случае воспроизведения громкоговорителями, декодер действует в режиме 103, в котором базовый декодер 115 декодирует безэховое представление Y , а также параметры W_Z . Впоследствии этап 116 матрицирования применяет параметры W_Z в отношении безэхового представления Y для получения оценки, или приближения, представления Z для громкоговорителей. Позднее этот сигнал преобразовывается во временную область блоком 118 синтезирующих фильтров HCQMF и воспроизводится громкоговорителями 140.

Наконец, следует отметить, что система, показанная на фиг. 7, необязательно, может эксплуатироваться при отсутствии определения и передачи параметров W_Z . В этом режиме работы представление Z для громкоговорителей нельзя сгенерировать из безэхового представления Y . Однако по причине того, что определяются и передаются параметры W_E и/или W_F , из безэхового представления можно сгенерировать представление для наушников, содержащее составляющие ранних отражений и/или поздней реверберации из безэхового представления.

Многоканальное представление для громкоговорителей.

Специалистам в данной области следует принять во внимание, что первое представление потока воспроизведения, закодированное в кодере, может представлять собой многоканальное представление, например, окружающее или погружающее представление для громкоговорителей, такое как представление в форматах 5.1, 7.1, 7.1.4 и т.д. Обсужденные выше, например со ссылкой на фиг. 4, варианты осуществления изобретения, в которых второе представление потока воспроизведения представляет собой сте-

реофоническое представление, будут действовать аналогичным образом, хотя размер матриц будет откорректирован.

Например, в то время как для преобразования из одного стереофонического представления в другое стереофоническое представление достаточной является матрица параметров размером 2×2 , для преобразования пятиканального окружающего представления в стереофоническое представление требуется матрица размером 5×2 , а для преобразования из окружающего представления в формате 5.1 (пять каналов полной полосы пропускания и канал низкочастотных эффектов (low-frequency effects, LFE)) в стереофоническое представление - матрица размером 6×2 . Как следствие, количество дополнительной информации, необходимой для представления параметров преобразования, будет увеличиваться вместе с числом каналов в представлении для громкоговорителей, и соответствующим образом также будет увеличиваться вычислительная сложность процесса декодирования.

Во избежание или для минимизации этого возрастания вычислительной сложности при преобразовании первого представления с $M1$ каналов во второе представление с $M2$ каналов, где $M1 > M2$, например, когда окружающее или погружающее представление для громкоговорителей преобразовывается в бинауральное стереофоническое представление, перед определением параметров преобразования может являться преимущественным понижающее микширование первого представления в промежуточное представление. Например, окружающее представление в формате 5.1 может быть подвергнуто понижающему микшированию в стереофоническое представление для громкоговорителей в формате 2.0.

На фиг. 8а показан кодер 200, в котором звуковое содержимое 201 представляется блоком 202 представления в окружающее представление S для громкоговорителей формата 5.1, которое кодируется базовым кодером 203. Представление S в формате 5.1 также преобразовывается модулем 204 понижающего микширования в полученное в результате понижающего микширования промежуточное двухканальное (стереофоническое) представление Z . Например, левый канал, Z (ZL), может быть выражен как взвешенная сумма левого канала (SL), левого бокового канала (SLS), центрального канала (SC) и канала низкочастотных эффектов ($SLFE$) окружающего представления S в соответствии со следующим уравнением:

$$ZL = (SL + a \cdot SC + b \cdot SLS + c \cdot SLFE),$$

где a , b и c представляют собой соответствующие постоянные, например $a=b=\sqrt{0,5}=0,71$ и $c=0,5$.

Звуковое содержимое также вводится в блок 205 бинаурального представления, выполненный с возможностью представления безэхового бинаурального сигнала Y . Блок 206 вычисления параметров принимает безэховый сигнал Y и стереофонический сигнал Z , полученный в результате понижающего микширования, и вычисляет параметры W_Y преобразования стереофонического сигнала в безэховый сигнал. В сравнении с вышеописанной фиг. 4, блок 202 представления представляет собой многоканальный вариант блока 45 представления, так как выходные данные в обоих случаях доставляются в базовый кодер 203/48. Блоки 205 и 206 в принципе идентичны блокам 43 и 46.

Кроме того, кодер также может содержать блок 207 (соответствующий блоку 82, показанному на фиг. 6) для представления входного сигнала FDN , и тогда блок 206 может быть выполнен с возможностью вычисления также и набора параметров W_F FDN (соответствующих блоку 83 на фиг. 6).

На фиг. 8b показан декодер 210, в котором базовый декодер 211 принимает и декодирует окружающее представление S в формате 5.1, а также наборы параметров W_Y и W_F . Окружающее представление S преобразовывается в полученный в результате понижающего микширования двухканальный (стереофонический) сигнал Z посредством модуля 212 понижающего микширования, действующего таким же образом, как его эквивалент 204 в кодере. Первый блок 213 матрицирования применяет параметры W_Y к стереофоническому представлению Z для создания реконструированного безэхового сигнала \hat{Y} .

Второй блок 214 матрицирования применяет параметры.

W_F к стереофоническому представлению Z для создания реконструированного входного сигнала FDN . Входной сигнал FDN используется в FDN 215 для создания сигнала поздней реверберации, который добавляется 216 к реконструированному безэховому сигналу \hat{Y} для создания бинауральных выходных данных. Следует отметить, что обработка в блоках 213-216 аналогична таковой для декодера 86, показанного на фиг. 6.

Для менее высоких битовых скоростей передачи данных известно использование параметрических способов передачи представления в формате 5.1 при помощи полученного в результате понижающего микширования сигнала в формате 2.1 и набора параметров связи, см., например, публикацию ETSI TS 103 190-1 V1.2.1 (2015-06). В такой системе базовый декодер фактически выполняет повышающее микширование для создания декодированного представления в формате 5.1. Если вариант осуществления, показанный на фиг. 8b, реализуется в таком декодере, результатом будет декодер, показанный на фиг. 9а. Следует отметить, что базовый декодер 311, показанный на фиг. 9а, содержит модуль 312 повышающего микширования для повышающего микширования представления в формате 2.1 в представление в формате 5.1. Представление в формате 5.1 затем подвергается понижающему микшированию в представление в формате 2.0 при помощи модуля 212 понижающего микширования, как показано на фиг. 8b.

Однако в данном контексте, если представление в формате 2.1 уже включено в битовый поток, повышающее микширование в формат 5.1 не требуется и может быть опущено для упрощения декодера. Такой упрощенный декодер показан на фиг. 9b. Здесь базовый декодер 411 декодирует только представление в формате 2.1. Это представление принимается упрощенным модулем 412 понижающего микширования, выполненным с возможностью преобразования представления в формате 2.1 в представление в формате 2.0 в соответствии с уравнениями:

$$\begin{aligned}L_o &= a*L + b*LFE; \\R_o &= a*R + b*LFE,\end{aligned}$$

где L, R и LFE обозначают левый и правый каналы полной полосы пропускания и канал низкочастотных эффектов декодированного представления в формате 2.1, при этом a и b представляют собой соответствующие постоянные, оказывающие воздействие на повышающее и понижающее микширование, выполняемое модулями 312 и 212, показанными на фиг. 9a.

Процесс, описанный на фиг. 9a и 9b, предполагает получение в результате понижающего микширования сигнала в формате 2.1 и соответствующих параметров связи. Аналогичный подход может быть использован в системе, в которой используется, например, полученный в результате понижающего микширования сигнал в формате 3.1 и соответствующие параметры связи. В качестве альтернативы система, показанная на фиг. 8a и 8b, также может переносить дополнительную вспомогательную информацию, позволяющую выполнять повышающее микширование представления в формате 5.1 в представление на основе объектов, как обсуждено в публикации ETSI TS 103 190-1 V1.2.1 (2015-06).

Толкования.

Отсылка в данном описании к "одному варианту осуществления", "некоторым вариантам осуществления" или "одному из вариантов осуществления" означает, что конкретный отличительный признак, конструкция или характеристика, описанная в связи с данным вариантом осуществления, включена по меньшей мере в один вариант осуществления настоящего изобретения. Поэтому появления фразы "в одном варианте осуществления", "в некоторых вариантах осуществления" или "в одном из вариантов осуществления" в различных местах данного описания могут, но необязательно, относиться к одному и тому же варианту осуществления изобретения. Кроме того, конкретные отличительные признаки, конструкции или характеристики могут комбинироваться в одном или более вариантах осуществления любым подходящим образом, что должно быть очевидно из данного описания для специалистов средней квалификации в данной области.

В рамках данного описания использование порядковых числительных "первый", "второй", "третий" и т.д. для описания обычного объекта указывает единственно на то, что производится отсылка к различным примерам сходных объектов, и они не предназначены для обозначения того, что объекты, описанные таким образом, должны находиться в данной последовательности во времени, в пространстве, по рангу или любым иным образом.

В приведенной ниже формуле изобретения и в данном описании любой из терминов "содержащий", "состоящий из" или "который содержит" является неограничивающим термином, что означает включение, по меньшей мере, следующих за ним элементов/отличительных признаков, но не исключение остальных. Поэтому термин "содержащий" при его использовании в формуле изобретения не следует интерпретировать как ограничивающий в отношении средств или элементов, или этапов, перечисляемых после него. Например, объем выражения "устройство, содержащее A и B" не следует ограничивать устройствами, содержащими только элементы A и B. Если любой из используемых в данном описании терминов "включающий" или "который включает" также представляет собой неограничивающий термин, который также означает включение, по меньшей мере, элементов/отличительных признаков, следующих за этим термином, но не исключение остальных. Таким образом, "включающий" является синонимом и означает "содержащий".

В рамках данного описания термин "примерный" используется в смысле представления примеров, в отличие от указания свойства. Т.е. "примерный вариант осуществления" - это вариант осуществления, предусмотренный в качестве примера, но не обязательно являющийся одним из вариантов осуществления примерного свойства.

Следует понимать, что в приведенном выше описании примерных вариантов осуществления изобретения различные отличительные признаки изобретения иногда группируются в один вариант осуществления изобретения, фигуру или их описание для выбора оптимального пути описания и для обеспечения понимания одного или более различных аспектов изобретения. Такой способ раскрытия, однако, не следует интерпретировать как отражение намерения того, что заявленное изобретение требует большего количества отличительных признаков, чем количество признаков, которые в прямой форме перечислены в каждом пункте формулы изобретения. Вместо этого, как отражает нижеследующая формула изобретения, особенности изобретения заключаются менее чем во всех отличительных признаках вышеописанного одного варианта осуществления изобретения. Поэтому формула изобретения, следующая за разделом "Подробное описание", таким образом безоговорочно включена в этот раздел "Подробное описание", причем каждый пункт формулы самостоятельно представляет собой отдельный вариант осуществления данного изобретения.

Кроме того, несмотря на то что некоторые варианты осуществления изобретения, описанные в данном описании, включают одни, а не другие отличительные признаки, включенные в другие варианты осуществления изобретения, комбинации отличительных признаков из различных вариантов осуществления изобретения подразумеваются как находящиеся в пределах объема изобретения и образующие другие варианты осуществления изобретения, как должно быть понятно специалистам в данной области. Например, в нижеприведенной формуле изобретения любые заявленные варианты осуществления изобретения могут быть использованы в любой комбинации.

Кроме того, некоторые варианты осуществления изобретения описаны в данном описании как способ или комбинация элементов способа, которые могут быть реализованы процессором вычислительной системы, или другими средствами осуществления такой функции. Таким образом, процессор с необходимыми командами для осуществления указанного способа или элемента способа образует средства для осуществления способа или элемента способа. Кроме того, описанный в данном описании элемент варианта осуществления устройства представляет собой пример средств осуществления функции, выполняемой элементом для осуществления изобретения.

В приведенном в данном документе описании изложено множество конкретных деталей. Однако следует понимать, что варианты осуществления изобретения могут использоваться на практике без этих конкретных деталей. В других случаях хорошо известные способы, конструкции и технологии подробно не показаны для того, чтобы не делать менее ясным понимание данного описания.

Аналогично, следует обратить внимание, что термин "связанный" при его использовании в формуле изобретения не следует интерпретировать как ограничивающийся только прямыми соединениями. Могут использоваться термины "связанный" и "соединенный" наряду с их производными. Следует понимать, что эти термины не предполагаются как синонимы друг друга. Таким образом, объем выражения "устройство А, связанное с устройством В" не следует ограничивать устройствами или системами, в которых вывод устройства А непосредственно соединен с вводом устройства В. Это означает, что существует путь между выводом устройства А и вводом устройства В, который может представлять собой путь, содержащий другие устройства или средства. "Связанный" может означать то, что два или большее количество элементов находятся или в прямом физическом, или электрическом контакте, или то, что два или большее количество элементов не находятся в прямом контакте друг с другом, однако по-прежнему кооперируются или взаимодействуют друг с другом.

Таким образом, в то время как здесь описано то, что рассматривается как предпочтительные варианты осуществления изобретения, специалистам в данной области должно быть понятно, что в них могут вноситься другие и дополнительные модификации без отступления от идеи изобретения, и подразумевается, что все указанные изменения и модификации заявляются как включенные в объем изобретения. Например, любые приведенные выше формулы являются только примерами процедур, которые могут использоваться. Функциональные возможности могут добавляться к структурным схемам или исключаться из структурных схем, а операции могут быть подвержены взаимному обмену между функциональными блоками. Этапы могут добавляться к способам или исключаться из способов, описанных в пределах объема настоящего изобретения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ кодирования входного аудиопотока, содержащего одну или более звуковых составляющих, отличающийся тем, что каждая звуковая составляющая связана с пространственным местоположением, причем способ включает этапы:

представление первого представления потока воспроизведения указанного входного аудиопотока, причем указанное первое представление потока воспроизведения представляет собой набор из $M1$ сигналов, предназначенных для воспроизведения первой системой воспроизведения звука;

представление второго представления потока воспроизведения указанного входного аудиопотока, при этом указанное второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление;

определение набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом набор параметров преобразования образует матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, и причем промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения;

причем параметры преобразования определяются путем минимизации величины разницы между вторым представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представлению потока воспроизведения; и

кодирование первого представления потока воспроизведения и указанного набора параметров преобразования для передачи в декодер.

2. Способ по п.1, отличающийся тем, что указанные параметры преобразования являются переменными во времени и/или зависящими от частоты.

3. Способ по любому из предыдущих пунктов, отличающийся тем, что $M1=2$.

4. Способ по любому из пп.1, 2, отличающийся тем, что $M1>2$, и при этом способ дополнительно включает образование промежуточного представления потока воспроизведения путем понижающего микширования первого представления потока воспроизведения в двухканальное представление.

5. Способ декодирования представлений потока воспроизведения из потока данных, причем способ включает этапы:

прием и декодирование первого представленного представления потока воспроизведения, при этом указанное первое представленное представление потока воспроизведения представляет собой набор из $M1$ сигналов, предназначенных для воспроизведения первой системой воспроизведения звука;

прием и декодирование набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представленного представления потока воспроизведения, при этом указанное второе представленное представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом набор параметров преобразования образуют матрицу усиления размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения;

при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление, и причем промежуточное представление потока воспроизведения представляет собой одно из первого представленного представления потока воспроизведения и результата понижающего микширования первого представленного представления потока воспроизведения;

причем параметры преобразования определяются кодером для обеспечения того, что величина разницы между вторым представленным представлением потока воспроизведения и приближением второго представленного представления потока воспроизведения, образованного применением параметров преобразования к первому представленному представлению потока воспроизведения, является минимальной; и

применение указанных параметров преобразования к указанному промежуточному представлению потока воспроизведения для получения указанного приближения второго представленного представления потока воспроизведения.

6. Способ по п.5, отличающийся тем, что указанные параметры преобразования являются переменными во времени и/или зависящими от частоты.

7. Способ по п.5, отличающийся тем, что $M1=2$.

8. Способ по любому из пп.5, 6, отличающийся тем, что $M1>2$, и при этом способ дополнительно включает образование промежуточного представления потока воспроизведения путем понижающего микширования первого представленного представления потока воспроизведения в двухканальное представление.

9. Способ по п.8, отличающийся тем, что первое представленное представление потока воспроизведения представляет собой окружающее или погружающее представление, такое как представление в формате 5.1, 7.1 или 7.1.4.

10. Способ по п.8, отличающийся тем, что первое представленное представление потока воспроизведения представляет собой представление в формате 2.1, и при этом указанный этап понижающего микширования включает преобразование представления в формате 2.1 в стереофоническое представление в формате 2.0 в соответствии с уравнениями:

$$\begin{aligned} L_o &= a*L + b*LFE, \\ R_o &= a*R + b*LFE, \end{aligned}$$

где L, R и LFE обозначают левый и правый каналы полной полосы пропускания и канал низкочастотных эффектов декодированного представления в формате 2.1, при этом a и b представляют собой соответствующие постоянные.

11. Способ по п.5, отличающийся тем, что приближение второго представленного представления потока воспроизведения представляет собой безэховое бинауральное представление, и при этом способ дополнительно включает:

прием и декодирование одного или более дополнительных наборов параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в один или более входных сигналов процесса имитирования акустической среды;

применение одного или более дополнительных наборов параметров преобразования к промежуточному представлению потока воспроизведения для генерирования одного или более входных сигналов процесса имитирования акустической среды;

применение одного или более входных сигналов процесса имитирования акустической среды к одному или более процессам имитирования акустической среды для получения одного или более сигналов имитированной акустической среды;

комбинирование одного или более сигналов имитированной акустической среды с приближением второго представленного представления потока воспроизведения.

12. Способ по п.5, отличающийся тем, что он дополнительно включает

прием и декодирование одного или более наборов параметров оценки диалоговых сигналов, пригодных для преобразования промежуточного представления потока воспроизведения в один или более оценочных диалоговых сигналов;

применение одного или более наборов параметров оценки диалога к промежуточному представлению потока воспроизведения для генерирования одного или более оценочных диалоговых сигналов;

вычитание одного или более оценочных диалоговых сигналов из промежуточного представления потока воспроизведения для получения одного или более сигналов с подавленным диалогом;

применение одного или более сигналов с подавленным диалогом к одному или более процессам имитирования акустической среды для получения одного или более сигналов имитированной акустической среды;

комбинирование одного или более сигналов имитированной акустической среды с приближением второго представленного представления потока воспроизведения.

13. Кодер для кодирования входного аудиопотока, содержащего одну или более звуковых составляющих, причем каждая звуковая составляющая связана с пространственным местоположением, причем кодер содержит

первый блок представления для представления первого представления потока воспроизведения указанного входного аудиопотока, при этом указанное первое представление потока воспроизведения представляет собой набор из M_1 сигналов, предназначенных для воспроизведения первой системой воспроизведения звука;

второй блок представления для представления второго представления потока воспроизведения указанного входного аудиопотока, при этом указанное второе представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представления потока воспроизведения и второго представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление;

блок определения параметров преобразования для определения набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представления потока воспроизведения, при этом набор параметров преобразования образует матрицу усиления размером $M_1 \times M_2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, причем промежуточное представление потока воспроизведения представляет собой одно из первого представления потока воспроизведения и результата понижающего микширования первого представления потока воспроизведения, и причем параметры преобразования определяются путем минимизации величины разницы между вторым представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представлению потока воспроизведения; и

блок кодирования для кодирования первого представления потока воспроизведения и указанного набора параметров преобразования для передачи в декодер.

14. Декодер для декодирования представлений потока воспроизведения из потока данных, причем декодер содержит:

блок базового декодера, выполненный с возможностью

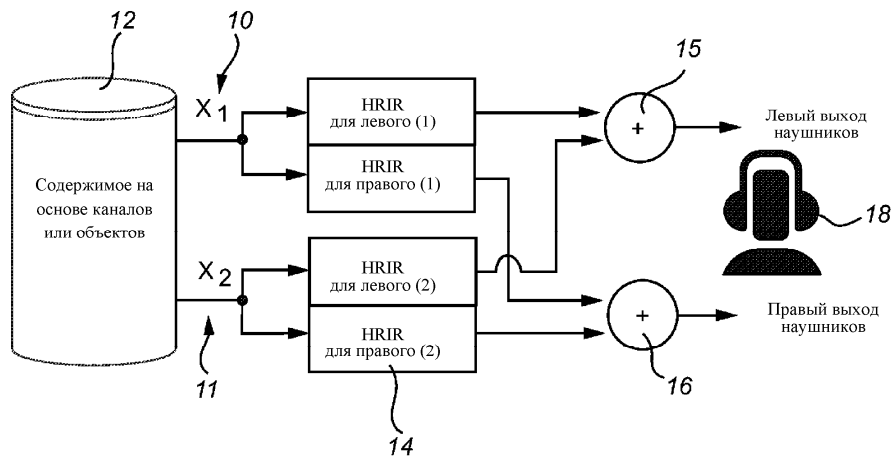
приема и декодирования первого представленного представления потока воспроизведения, при этом указанное первое представленное представление потока воспроизведения представляет собой набор из M_1 сигналов, предназначенных для воспроизведения первой системой воспроизведения звука, и

приема и декодирования набора параметров преобразования, пригодных для преобразования промежуточного представления потока воспроизведения в приближение второго представленного представления потока воспроизведения, при этом указанное второе представленное представление потока воспроизведения представляет собой набор из двух сигналов, предназначенных для воспроизведения второй системой воспроизведения звука, при этом набор параметров преобразования образует матрицу усиления

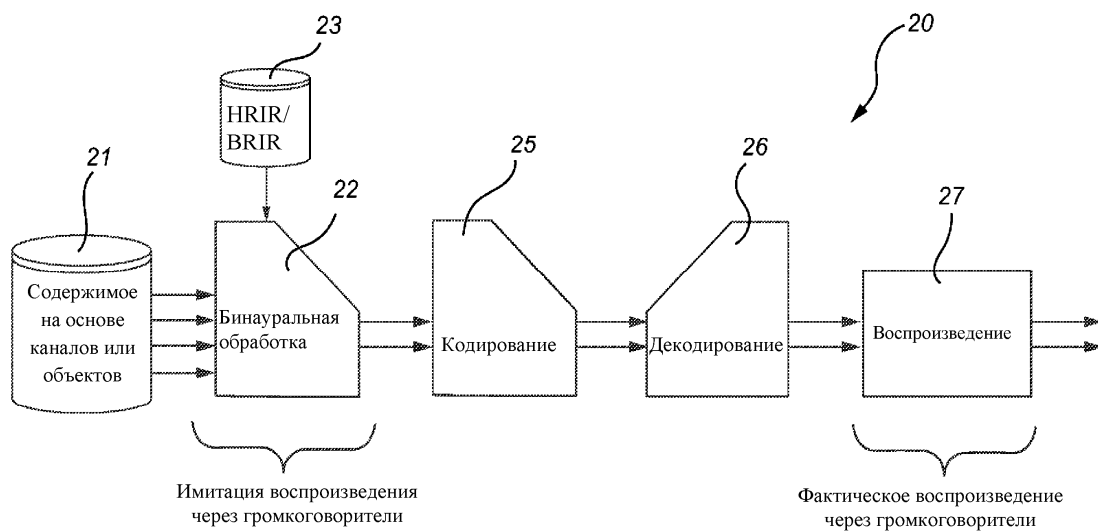
размером $M1 \times M2$, которая при применении непосредственно к промежуточному представлению потока воспроизведения матричным умножением образует указанное приближение второго представления потока воспроизведения, при этом одно из первого представленного представления потока воспроизведения и второго представленного представления потока воспроизведения представляет собой представление для громкоговорителей, и при этом одно из первого представленного представления потока воспроизведения и второго представленного представления потока воспроизведения представляет собой эховое или безэховое бинауральное представление, и причем промежуточное представление потока воспроизведения представляет собой одно из первого представленного представления потока воспроизведения и результата понижающего микширования первого представленного представления потока воспроизведения, причем параметры преобразования определяются кодером для обеспечения того, что величина разницы между вторым представленным представлением потока воспроизведения и приближением второго представления потока воспроизведения, образованного применением параметров преобразования к первому представленному представлению потока воспроизведения, является минимальной; и

матричный умножитель для применения указанных параметров преобразования к указанному промежуточному представлению потока воспроизведения для получения указанного приближения второго представленного представления потока воспроизведения.

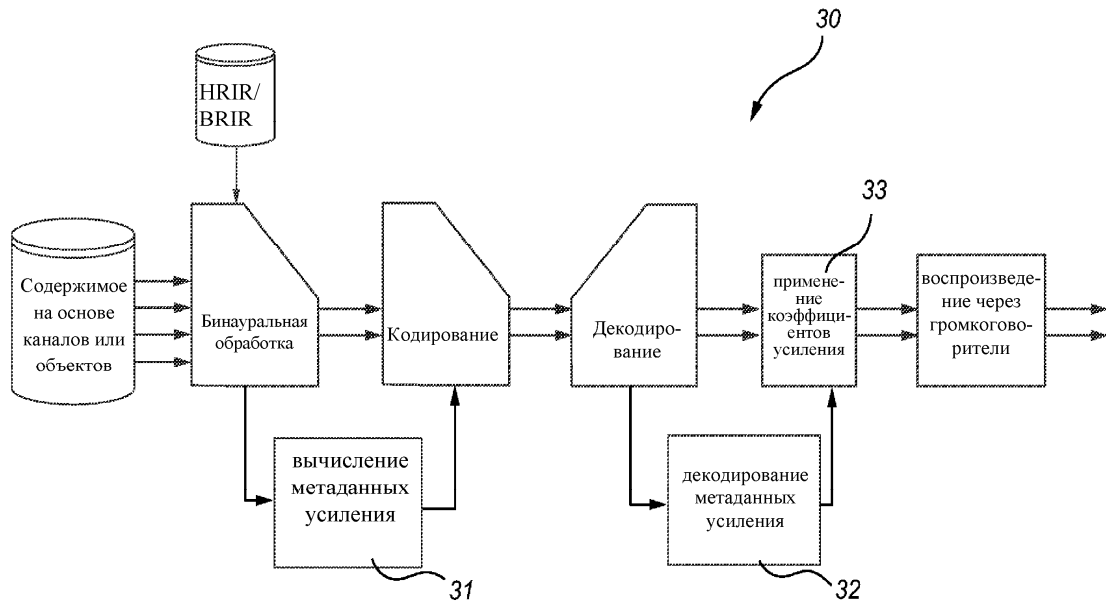
15. Внутренний носитель данных, на котором хранится компьютерный программный продукт кодера, содержащий сегменты компьютерного программного кода, которые при их исполнении в компьютерном процессоре вызывают выполнение компьютерным процессором способа по любому из пп.1-12.



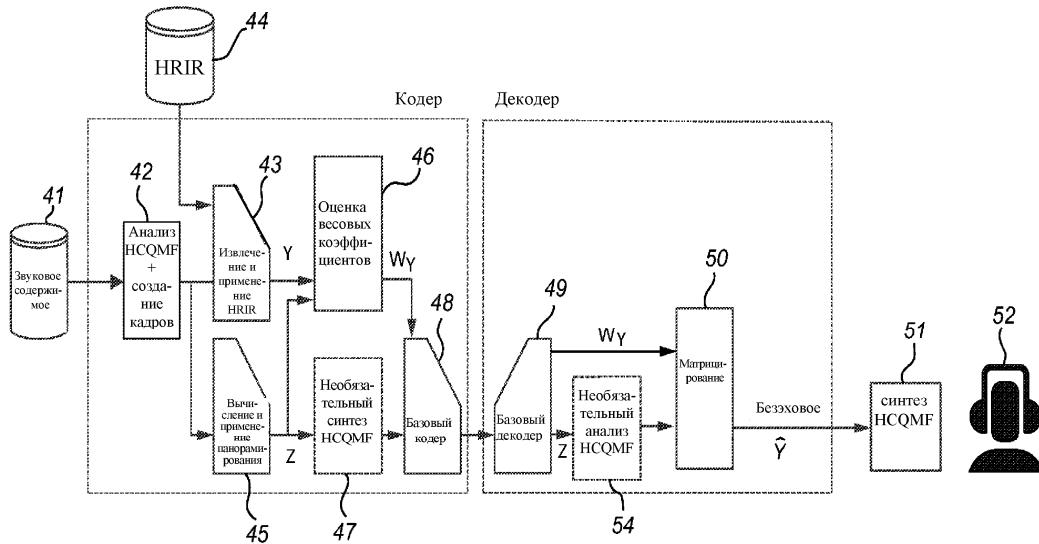
Фиг. 1



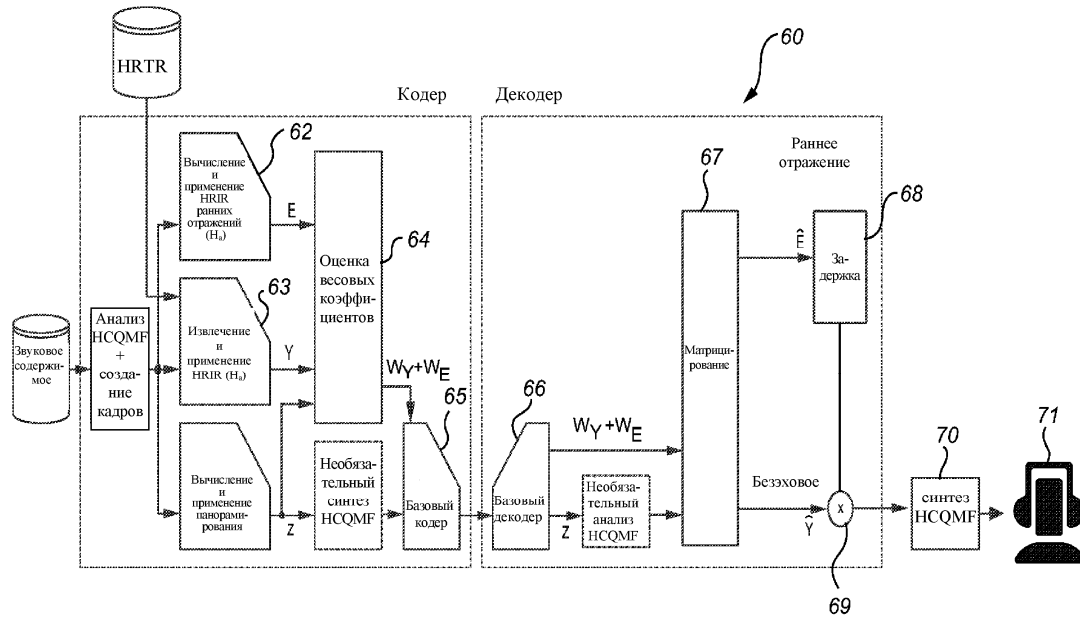
Фиг. 2 (Уровень техники)



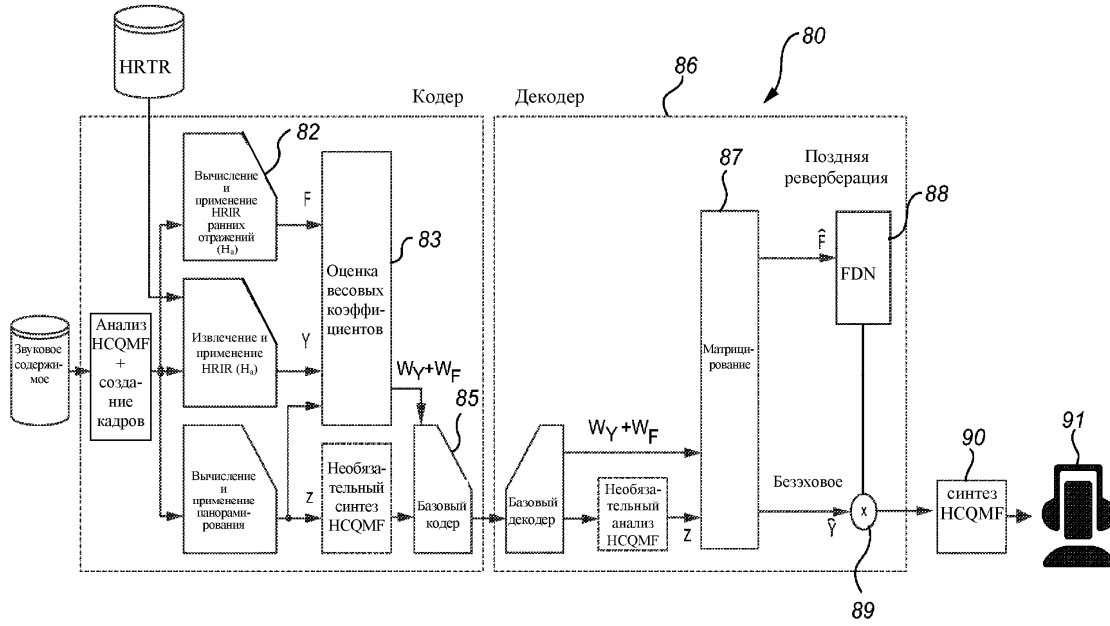
Фиг. 3



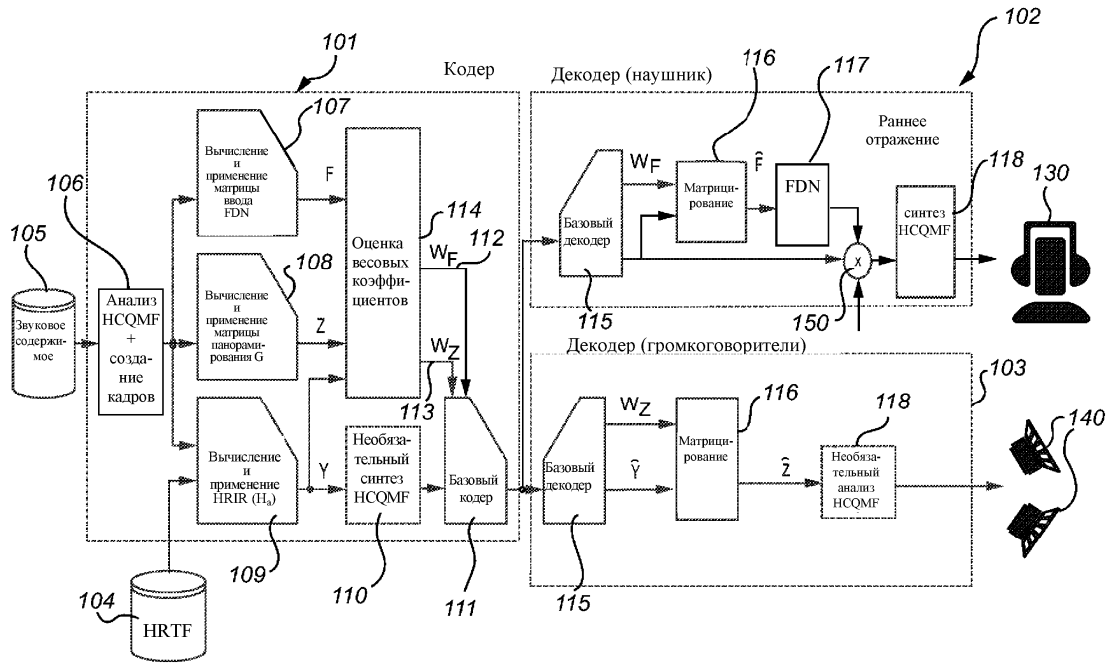
Фиг. 4



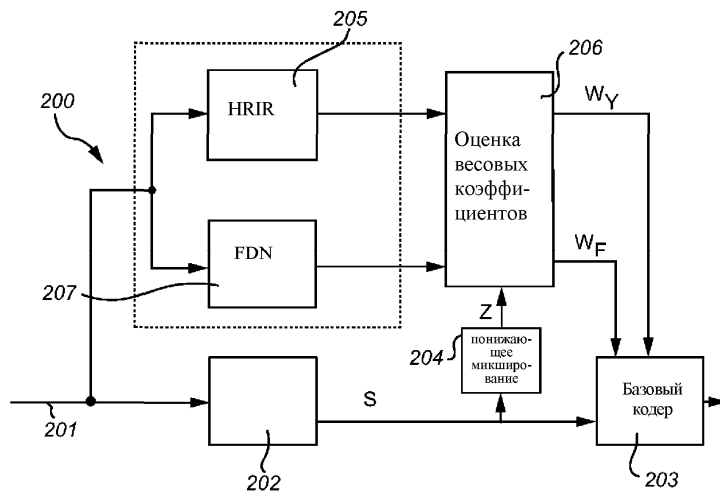
Фиг. 5



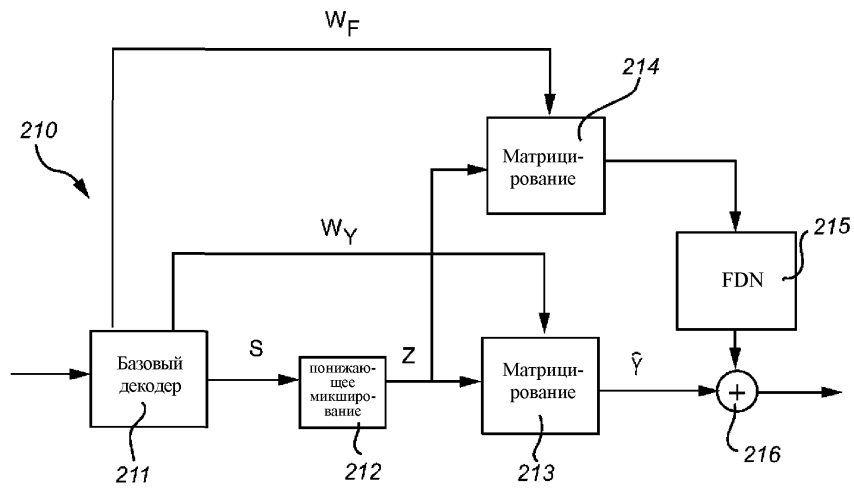
Фиг. 6



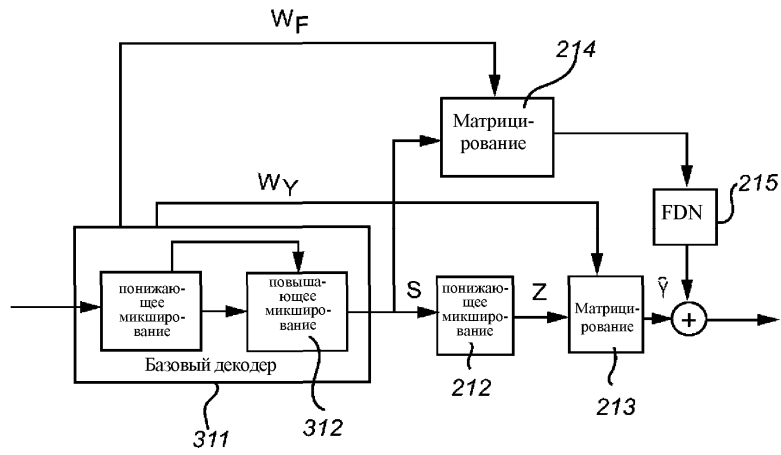
Фиг. 7



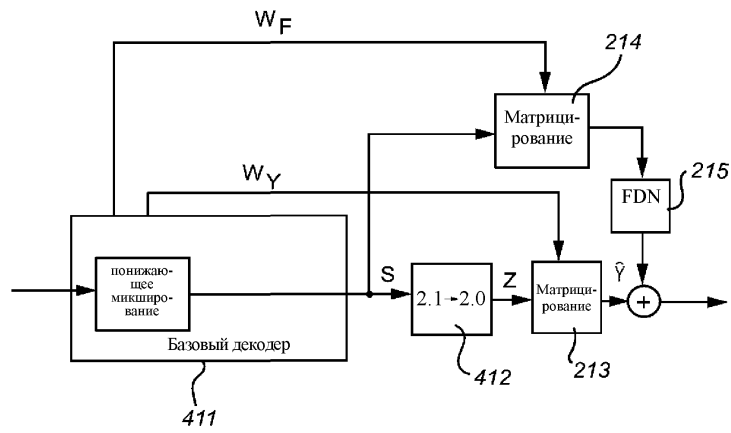
Фиг. 8a



Фиг. 8b



Фиг. 9а



Фиг. 9b

