

(19)



Евразийское  
патентное  
ведомство

(21) 201890843 (13) A1

## (12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки  
2018.10.31

(51) Int. Cl. G10L 19/008 (2013.01)

(22) Дата подачи заявки  
2016.10.07

## (54) МНОГОУРОВНЕВОЕ КОДИРОВАНИЕ СЖАТЫХ ПРЕДСТАВЛЕНИЙ ЗВУКА ИЛИ ЗВУКОВОГО ПОЛЯ

(31) 15306589.1; 15306653.5; 62/361,461;  
62/361,416(32) 2015.10.08; 2015.10.15; 2016.07.12;  
2016.07.12

(33) EP; EP; US; US

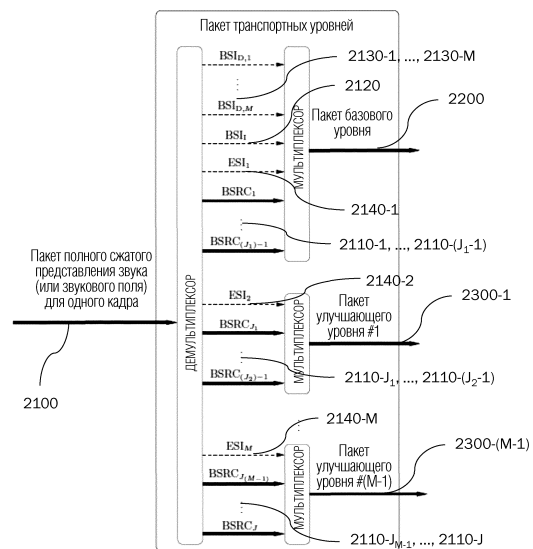
(86) PCT/EP2016/073969

(87) WO 2017/060410 2017.04.13

(71) Заявитель:  
ДОЛБИ ИНТЕРНЭШНЛ АБ (NL)(72) Изобретатель:  
Кордон Свен, Крюгер Александр (DE)(74) Представитель:  
Медведев В.Н. (RU)

(57) Настоящий документ относится к способу многоуровневого кодирования сжатого представления звука или звукового поля. Сжатое представление звука содержит базовое сжатое представление звука, содержащее множество компонентов, базовую вспомогательную информацию для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля и улучшающую вспомогательную информацию, включающую в себя параметры для улучшения базового воссозданного представления звука. Способ содержит подразделение множества компонентов на множество групп компонентов и присвоение каждой из множества групп соответствующему одному из множества иерархических уровней, количество групп соответствует количеству уровней, и множество уровней включает в себя базовый уровень и один или более иерархических улучшающих уровней, добавление базовой вспомогательной информации к базовому уровню и определение множества частей улучша-

ющей вспомогательной информации из улучшающей вспомогательной информации и присвоение каждой из множества частей улучшающей вспомогательной информации соответствующему одному из множества уровней, причем каждая часть улучшающей вспомогательной информации включает в себя параметры для улучшения воссозданного представления звука, доступные из данных, включенных в соответствующий уровень и любые уровни ниже соответствующего уровня. Документ также относится к способу декодирования сжатого представления звука или звукового поля, причем сжатое представление звука закодировано в множестве иерархических уровней, которые включают в себя базовый уровень и один или более иерархических улучшающих уровней, а также к кодеру и декодеру для многоуровневого кодирования сжатого представления звука.



A1

201890843

201890843

A1

## ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-549282EA/019

### МНОГОУРОВНЕВОЕ КОДИРОВАНИЕ СЖАТЫХ ПРЕДСТАВЛЕНИЙ ЗВУКА ИЛИ ЗВУКОВОГО ПОЛЯ

#### ПЕРЕКРЕСТНАЯ ССЫЛКА НА РОДСТВЕННЫЕ ЗАЯВКИ

По настоящей заявке испрашивается приоритет по европейским патентным заявкам № 15306589.1, поданной 8 октября 2015 года, и № 15306653.5, поданной 15 октября 2015 года, и патентным заявкам США № 62/361,461 и № 62/361,416, содержание которых полностью включено в настоящую заявку посредством ссылки.

#### ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

Настоящий документ относится к способам и устройствам для многоуровневого аудиокодирования. В частности, настоящий документ относится к способам и устройствам для многоуровневого аудиокодирования сжатых представлений звука (или звукового поля), например, представления звука (или звукового поля) системы Амбисоник высшего порядка (Higher-Order Ambisonics, HOA).

#### УРОВЕНЬ ТЕХНИКИ

Для потокового представления звука (или звукового поля) по каналу передачи с изменяющимися по времени условиями многоуровневое кодирование является средством, чтобы адаптировать качество принятого представления звука к условиям передачи и, в частности, избежать нежелательных пропаданий сигнала.

Для многоуровневого кодирования представление звука (или звукового поля) обычно подразделяется на высокоприоритетный базовый уровень относительно небольшого размера и дополнительные улучшающие уровни с убывающими приоритетами и произвольными размерами. Каждый улучшающий уровень, как обычно предполагается, содержит нарастающую информацию, чтобы дополнить все из более низких уровней для улучшения качества представления звука (или звукового поля). Величиной защиты от ошибок для передачи отдельных уровней управляют на основе их приоритета. В частности, базовому уровню предоставляется высокая защита от ошибок, которая является разумной и приемлемой вследствие ее

малого размера.

Однако существует потребность в многоуровневых схемах кодирования для (расширенной версии) специальных типов сжатых представлений звука или звуковых полей, таких как, например, сжатые представления НОА звука или звукового поля.

Настоящий документ решает упомянутые выше проблемы. В частности, описаны способы и кодеры/декодеры для многоуровневого кодирования сжатых представлений звука или звукового поля описаны.

#### СУЩНОСТЬ ИЗОБРЕТЕНИЯ

В соответствии с аспектом описан способ многоуровневого кодирования сжатого представления звука или звукового поля. Сжатое представление звука может включать в себя базовое сжатое представление звука, которое включает в себя множество компонентов. Множество компонентов могут являться взаимодополняющими компонентами. Сжатое представление звука может дополнительно включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля. Сжатое представление звука может, кроме того, включать в себя улучшающую вспомогательную информацию, включающую в себя параметры для улучшения (например, расширения) базового воссозданного представления звука. Способ может включать в себя подразделение (например, группировку) множества компонентов на множество групп компонентов. Способ может дополнительно включать в себя присвоение (например, добавление) каждой из множества групп соответствующему одному из множества иерархических уровней. Присвоение может указывать соответствие между соответствующими группами и уровнями. Можно сказать, что компоненты, присвоенные соответствующему уровню, включены в этот уровень. Количество групп может соответствовать (например, быть равным) количеству уровней. Множество уровней может включать в себя базовый уровень и один или более иерархических улучшающих уровней. Множество иерархических уровней может быть упорядочено от базового уровня, через первый улучшающий уровень, второй улучшающий уровень и т.д., вплоть до общего наиболее высокого

улучшающего уровня (общего наиболее высокого уровня). Способ может дополнительно включать в себя добавление базовой вспомогательной информации к базовому уровню (например, включение базовой вспомогательной информации в базовый уровень или распределение базовой вспомогательной информации базовому уровню, например, в целях передачи или хранения). Способ может дополнительно включать в себя определение множества частей улучшающей вспомогательной информации на основе улучшающей вспомогательной информации. Способ может, кроме того, включать в себя присвоение (например, добавление) каждой из множества частей улучшающей вспомогательной информации соответствующему одному из множества уровней. Каждая часть улучшающей вспомогательной информации может включать в себя параметры для улучшения воссозданного (например, восстановленного) представления звука, доступные из данных, включенных (например, присвоенных или добавленных) в соответствующий уровень и любые уровни ниже соответствующего уровня. Многоуровневое кодирование может быть выполнено в целях передачи по каналу передачи или в целях сохранения на подходящем запоминающем носителе, таком как, например, CD, DVD или Blu-ray Disc™.

Сконфигурированный, как упомянуто выше, предложенный способ позволяет эффективно применять многоуровневое кодирование к сжатым представлениям звука, содержащим множество компонентов, а также базовую и улучшающую вспомогательную информацию (например, независимую базовую вспомогательную информацию и улучшающую вспомогательную информацию), имеющие изложенные выше свойства. В частности, предложенный способ гарантирует, что каждый уровень включает в себя подходящую вспомогательную информацию для восстановления воссозданного представления звука из компонентов, включенных в любые уровни вплоть до рассматриваемого уровня. При этом предполагается, что уровни вплоть до рассматриваемого уровня включают в себя, например, базовый уровень, первый улучшающий уровень, второй улучшающий уровень и т.д., вплоть до рассматриваемого уровня. Таким образом, независимо от фактического наиболее высокого применимого уровня (например,

уровня ниже наиболее низкого слоя, который не был корректно принят, и, таким образом, все уровни ниже наиболее высокого применимого уровня и сам наиболее высокий применимый уровень приняты корректно), декодеру позволяется улучшить или расширить воссозданное представление звука, даже если воссозданное представление звука может отличаться от полного представления звука. В частности, независимо от фактического наиболее высокого применимого уровня, для декодера достаточно декодировать полезную нагрузку улучшающей вспомогательной информации только для единственного слоя (т.е., для наиболее высокого применимого уровня), чтобы улучшить или расширить воссозданное представление звука, которое доступно на основе всех компонентов, включенных в уровни вплоть до фактического наиболее высокого применимого уровня. Таким образом, для каждого временного интервала (например, кадра) должна быть декодирована только единственная полезная нагрузка улучшающей вспомогательной информации. С другой стороны, предложенный способ позволяет полностью использовать преимущество сокращения требуемой ширины полосы, которое может быть достигнуто при применении многоуровневого кодирования.

В вариантах осуществления компоненты базового сжатого представления звука могут соответствовать монауральным сигналам (например, транспортным сигналам или монауральным транспортным сигналам). Монауральные сигналы могут представлять либо преобладающие звуковые сигналы, либо последовательности коэффициентов представления НОА. Монауральные сигналы могут быть квантованы.

В вариантах осуществления базовая вспомогательная информация может включать в себя информацию, которая определяет декодирование (например, восстановление) одного или более из множества компонентов индивидуально, независимо от других компонентов. Например, базовая вспомогательная информация может представлять вспомогательную информацию, относящуюся к индивидуальным монауральным сигналам, независимо от других монауральных сигналов. Таким образом, базовая вспомогательная информация может упоминаться как независимая базовая

вспомогательная информация.

В вариантах осуществления улучшающая вспомогательная информация может представлять улучшающую вспомогательную информацию. Улучшающая вспомогательная информация может включать в себя параметры предсказания для базового сжатого представления звука для улучшения (например, расширения) базового воссозданного представления звука, которые доступны из базового сжатого представления звука и базовой вспомогательной информации.

В вариантах осуществления способ может дополнительно включать в себя формирование транспортного потока для передачи данных множества уровней (например, данных, присвоенных или добавленных к соответствующим уровням, или иным образом включенных в соответствующие уровни). Базовый уровень может иметь наиболее высокий приоритет передачи, и иерархические улучшающие уровни могут иметь убывающие приоритеты передачи. Таким образом, приоритет передачи может уменьшаться от базового уровня до первого улучшающего уровня, от первого улучшающего уровня до второго улучшающего уровня и т.д. Величиной защиты от ошибок для передачи данных множества уровней можно управлять в соответствии с соответствующими приоритетами передачи. Тем самым может быть обеспечено, что по меньшей мере некоторое количество более низких уровней передаются достоверно, с другой стороны, сокращая полную требуемую ширину полосы без применения чрезмерной защиты от ошибок к более высоким уровням.

В вариантах осуществления способ может дополнительно включать в себя для каждого из множества уровней формирование пакета транспортного уровня, включающего в себя данные соответствующего уровня. Например, для каждого временного интервала (например, кадра) соответствующий пакет транспортного уровня может быть сформирован для каждого из множества уровней.

В вариантах осуществления сжатое представление звука может дополнительно включать в себя дополнительную базовую вспомогательную информацию для декодирования базового сжатого представления звука в базовое воссозданное представление звука. Дополнительная базовая вспомогательная информация может включать

в себя информацию, которая определяет декодирование одного или более из множества компонентов в зависимости от соответствующих других компонентов. Способ может дополнительно включать в себя выполнение декомпозиции дополнительной базовой вспомогательной информации на множество частей дополнительной базовой вспомогательной информации. Способ может, кроме того, включать в себя добавление частей дополнительной базовой вспомогательной информации к базовому уровню (например, включение части дополнительной базовой вспомогательной информации в базовый уровень или распределение частей дополнительной базовой вспомогательной информации базовому уровню, например в целях передачи или хранения). Каждая часть дополнительной базовой вспомогательной информации может быть связана с соответствующим уровнем и может включать в себя информацию, которая определяет декодирование одного или более компонентов, присвоенных соответствующему уровню, в зависимости (только) от соответствующих других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня. Таким образом, каждая часть дополнительной базовой вспомогательной информации определяет компоненты на соответствующем уровне, которому соответствует эта часть дополнительной базовой вспомогательной информации, независимо от любых других компонентов, присвоенных более высоким уровням, чем соответствующий уровень.

Сконфигурированный таким образом предложенный способ избегает фрагментации дополнительной базовой вспомогательной информации посредством добавления всех частей к базовому уровню. Другими словами, все части дополнительной базовой вспомогательной информации включены в базовый уровень. Декомпозиция дополнительной базовой вспомогательной информации гарантирует, что для каждого уровня доступна часть дополнительной базовой вспомогательной информации, что не требует знания компонентов на более высоких уровнях. Таким образом, независимо от фактического наиболее высокого применимого уровня, для декодера достаточно декодировать дополнительную базовую вспомогательную информацию, включенную в

уровни вплоть до наиболее высокого применимого уровня.

В вариантах осуществления дополнительная базовая вспомогательная информация может включать в себя информацию, которая определяет декодирование (например, восстановление) одного или более из множества компонентов в зависимости от других компонентов. Например, дополнительная базовая вспомогательная информация может представлять вспомогательную информацию, относящуюся к индивидуальным монауральным сигналам, в зависимости от других монауральных сигналов. Таким образом, дополнительная базовая вспомогательная информация может упоминаться как зависимая базовая вспомогательная информация.

В вариантах осуществления сжатое представление звука может быть обработано для последовательных временных интервалов, например, временных интервалов равного размера. Последовательные временные интервалы могут являться кадрами. Таким образом, способ может работать на основе кадров, т.е., сжатое представление звука может быть закодировано покадрово. Сжатое представление звука может быть доступно для каждого последовательного временного интервала (например, для каждого кадра). Таким образом, операция сжатия, посредством которой было получено сжатое представление звука, может работать на основе кадров.

В вариантах осуществления способ может дополнительно включать в себя формирование информации конфигурации, которая указывает для каждого уровня компоненты базового сжатого представления звука, которые присвоены этому уровню. Таким образом, декодер может без затруднений осуществить доступ к информации, необходимой для декодирования, без ненужного анализа принятых полезных нагрузок данных.

В соответствии с другим аспектом описан способ многоуровневого кодирования сжатого представления звука или звукового поля. Сжатое представление звука может включать в себя базовое сжатое представление звука, которое включает в себя множество компонентов. Множество компонентов может являться взаимодополняющими компонентами. Сжатое представление звука может дополнительно включать в себя базовую вспомогательную



информацию (например, независимую базовую вспомогательную информацию) и третью информацию (например, зависимую базовую вспомогательную информацию) для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля. Базовая вспомогательная информация может включать в себя информацию, которая определяет декодирование одного или более из множества компонентов индивидуально, независимо от других компонентов. Дополнительная базовая вспомогательная информация может включать в себя информацию, которая определяет декодирование одного или более из множества компонентов в зависимости от соответствующих других компонентов. Способ может включать в себя подразделение (например, группировку) множества компонентов на множество групп компонентов. Способ может дополнительно включать в себя присвоение (например, добавление) каждой из множества групп соответствующему одному из множества иерархических уровней. Присвоение может указывать соответствие между соответствующими группами и уровнями. Можно сказать, что компоненты, присвоенные соответствующему уровню, включены в тот уровень. Количество групп может соответствовать (например, быть равным) количеству уровней. Множество уровней может включать в себя базовый уровень и один или более иерархических улучшающих уровней. Способ может дополнительно включать в себя добавление базовой вспомогательной информации к базовому уровню (например, включение базовой вспомогательной информации в базовый уровень или распределение базовой вспомогательной информации базовому уровню, например в целях передачи или хранения). Способ может дополнительно включать в себя выполнение декомпозиции дополнительной базовой вспомогательной информации на множество частей дополнительной базовой вспомогательной информации и добавление частей дополнительной базовой вспомогательной информации к базовому уровню (например, включение частей дополнительной базовой вспомогательной информации в базовый уровень или распределение частей дополнительной базовой вспомогательной информации базовому уровню, например, в целях передачи или хранения). Каждая часть дополнительной базовой вспомогательной информации

может быть связана с соответствующим уровнем и включать в себя информацию, которая определяет декодирование одного или более компонентов, присвоенных соответствующему уровню, в зависимости от соответствующих других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня.

Сконфигурированный таим образом предложенный способ гарантирует, что для каждого уровня доступна подходящая дополнительная базовая вспомогательная информация для декодирования компонентов, включенных в любой уровень вплоть до соответствующего уровня, не требуя корректного приема или декодирования (или, в целом, знания) любых более высоких уровней. В случае сжатого представления НОА предложенный способ гарантирует, что в режиме векторного кодирования подходящий V-вектор доступен для всего компонента, принадлежащего уровням вплоть до наиболее высокого применимого уровня. В частности, предложенный способ исключает случай, в котором элементы V-вектора, соответствующего компонентам на более высоких уровнях, явно не сообщены. В соответствии с этим информация, включенная в уровни вплоть до наиболее высокого применимого уровня, является достаточной для декодирования (например, восстановления) любых компонентов, принадлежащих уровням вплоть до наиболее высокого применимого уровня. Тем самым обеспечивается подходящая восстановление соответствующих воссозданных представлений НОА для более низких уровней, даже если более высокие уровни не могли быть корректно приняты декодером. С другой стороны, предложенный способ позволяет полностью использовать преимущества сокращения требуемой ширины полосы, которое может быть достигнуто при применении многоуровневого кодирования.

Варианты осуществления этого аспекта могут относиться к вариантам осуществления предыдущего аспекта.

В соответствии с другим аспектом описан способ декодирования сжатого представления звука или звукового поля. Сжатое представление звука может быть закодировано во множестве иерархических уровней. Множество иерархических уровней может включать в себя базовый уровень и один или несколько

иерархических возрастающих уровней. Множество уровней может иметь присвоенные им компоненты базового сжатого представления звука или звукового поля. Другими словами, множество уровней может включать в себя компоненты базовой сжатой вспомогательной информации. Компоненты могут быть присвоены соответствующим уровням в соответствующих группах компонентов. Множество компонентов могут являться взаимодополняющими компонентами. Базовый уровень может включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука. Каждый уровень может включать в себя часть улучшающей вспомогательной информации, включающую в себя параметры для улучшения базового воссозданного представления звука, доступные из данных, включенных в соответствующий уровень и любые уровни ниже соответствующего уровня. Способ может включать в себя прием полезных нагрузок данных, соответственно связанных с множеством иерархических уровней. Способ может дополнительно включать в себя определение первого индекса уровня, указывающего наиболее высокий применимый уровень среди множества уровней для использования для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля. Способ может дополнительно включать в себя получение базового воссозданного представления звука из компонентов, присвоенных наиболее высокому применимому уровню и любым уровням ниже наиболее высокого применимого уровня, с использованием базовой вспомогательной информации. Способ может дополнительно включать в себя определение второго индекса уровня, который указывает, какая часть улучшающей вспомогательной информации должна использоваться для улучшения (например, расширения) базового воссозданного представления звука. Способ, кроме того, может включать в себя получение воссозданного представления звука или звукового поля из базового воссозданного представления звука со ссылкой на второй индекс уровня.

Сконфигурированный таким образом предложенный способ гарантирует, что воссозданное представление звука имеет оптимальное качество с использованием доступной (например, корректно принятой) информации в наилучшей возможной степени.

В вариантах осуществления компоненты базового сжатого представления звука могут соответствовать монауральным сигналам (например, монауральным транспортным сигналам). Монауральные сигналы могут представлять либо преобладающие звуковые сигналы, либо последовательности коэффициентов представления НОА. Монауральные сигналы могут быть квантованы.

В вариантах осуществления базовая вспомогательная информация может включать в себя информацию, которая определяет декодирование (например, восстановление) одного или более из множества компонентов индивидуально, независимо от других компонентов. Например, базовая вспомогательная информация может представлять вспомогательную информацию, относящуюся к индивидуальным монауральным сигналам, независимо от других монауральных сигналов. Таким образом, базовая вспомогательная информация может упоминаться как независимая базовая вспомогательная информация.

В вариантах осуществления улучшающая вспомогательная информация может представлять улучшающую вспомогательную информацию. Улучшающая вспомогательная информация может включать в себя параметры предсказания для базового сжатого представления звука для улучшения (например, расширения) базового воссозданного представления звука, которые доступны из базового сжатого представления звука и базовой вспомогательной информации.

В вариантах осуществления способ может дополнительно включать в себя определение для каждого уровня, был ли соответствующий уровень принят корректно. Способ может дополнительно включать в себя определение первого индекса уровня как индекса того уровня, который находится непосредственно ниже наиболее низкого уровня, который не был принят корректно.

В вариантах осуществления определение второго индекса уровня может включать в себя либо определение второго индекса уровня как равного первому индексу уровня, либо определению значения индекса в качестве второго индекса уровня, которое указывает, что не следует использовать какую-либо улучшающую вспомогательную информацию при получении воссозданного

представления звука. В последнем случае воссозданное представление звука может быть эквивалентно базовому воссозданному представлению звука.

В вариантах осуществления полезные нагрузки данных могут быть приняты и обработаны для последовательных временных интервалов, например временных интервалов равного размера. Последовательные временные интервалы могут являться кадрами. Таким образом, способ может работать на основе кадров. Способ может дополнительно включать в себя определение второго индекса уровня как равного первому индексу уровня, если сжатые представления звука для последовательных временных интервалов могут быть декодированы независимо друг от друга.

В вариантах осуществления полезные нагрузки данных могут быть приняты и обработаны для последовательных временных интервалов, например, временных интервалов равного размера. Последовательные временные интервалы могут являться кадрами. Таким образом, способ может работать на основе кадров. Способ может дополнительно включать в себя для данного временного интервала среди последовательных временных интервалов определение для каждого уровня, был ли соответствующий уровень принят корректно, если сжатые представления звука для последовательных временных интервалов не могут быть декодированы независимо друг от друга. Способ может дополнительно включать в себя определение первого индекса уровня для данного временного интервала как меньшего индекса из первого индекса уровня временного интервала, предшествующего данному временному интервалу, и индекса уровня, находящегося непосредственно ниже наиболее низкого уровня, который не был принят корректно.

В вариантах осуществления способ может дополнительно включать в себя для данного временного интервала определение, равен ли первый индекс уровня для данного временного интервала первому индексу уровня для предыдущего временного интервала, если сжатые представления звука для последовательных временных интервалов не могут быть декодированы независимо друг от друга. Способ может дополнительно включать в себя определение, что второй индекс уровня для данного временного интервала равен

первому индексу уровня для данного временного интервала, если первый индекс уровня для данного временного интервала равен первому индексу уровня для предыдущего временного интервала. Способ может дополнительно включать в себя определение значения индекса в качестве второго индекса уровня, которое указывает, что не следует использовать какую-либо улучшающую вспомогательную информацию при получении воссозданного представления звука, если первый индекс уровня для данного временного интервала не равен первому индексу уровня для предыдущего временного интервала.

В вариантах осуществления базовый уровень может включать в себя по меньшей мере одну часть дополнительной базовой вспомогательной информации, связанную с соответствующим уровнем и включающую в себя информацию, которая определяет декодирование одного или более компонентов среди компонентов, присвоенных соответствующему уровню, в зависимости от других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня. Способ может дополнительно включать в себя для каждой части дополнительной базовой вспомогательной информации декодирование части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные ее соответствующему уровню и любым уровням ниже соответствующего уровня. Способ может дополнительно включать в себя коррекцию части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные наиболее высокому применимому уровню и любым уровням между наиболее высоким применимым уровнем и соответствующим уровнем. Базовое воссозданное представление звука может быть получено из компонентов, присвоенных наиболее высокому применимому уровню и любым уровням ниже наиболее высокого применимого уровня, с использованием базовой вспомогательной информации и скорректированной части дополнительной базовой вспомогательной информации, полученной из частей дополнительной базовой вспомогательной информации, соответствующих уровням вплоть до наиболее высокого применимого уровня.

В вариантах осуществления дополнительная базовая

вспомогательная информация может включать в себя информацию, которая определяет декодирование (например, восстановление) одного или более из множества компонентов в зависимости от других компонентов. Например, дополнительная базовая вспомогательная информация может представлять вспомогательную информацию, относящуюся к индивидуальным монауральным сигналам, в зависимости от других монауральных сигналов. Таким образом, дополнительная базовая вспомогательная информация может упоминаться как зависимая базовая вспомогательная информация.

В соответствии с другим аспектом описан способ декодирования сжатого представления звука или звукового поля. Сжатое представление звука может быть закодировано во множестве иерархических уровней. Множество иерархических уровней может включать в себя базовый уровень и один или несколько иерархических возрастающих уровней. Множество уровней может иметь присвоенные им компоненты базового сжатого представления звука или звукового поля. Другими словами, множество уровней может включать в себя компоненты базовой сжатой вспомогательной информации. Компоненты могут быть присвоены соответствующим уровням в соответствующих группах компонентов. Множество компонентов могут являться взаимодополняющими компонентами. Базовый уровень может включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука. Базовый уровень может дополнительно включать в себя по меньшей мере одну часть дополнительной базовой вспомогательной информации, связанную с соответствующим уровнем и включающую в себя информацию, которая определяет декодирование одного или более компонентов среди компонентов, присвоенных соответствующему уровню в зависимости от других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня. Способ может включать в себя прием полезных нагрузок данных, соответственно связанных с множеством иерархических уровней. Способ может дополнительно включать в себя определение первого индекса уровня, указывающего наиболее высокий применимый уровень среди множества уровней для использования для декодирования базового сжатого представления

звука в базовое воссозданное представление звука или звукового поля. Способ может дополнительно включать в себя для каждой части дополнительной базовой вспомогательной информации декодирование части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные ее соответствующему уровню и любым уровням ниже соответствующего уровня. Способ может дополнительно включать в себя для каждой части дополнительной базовой вспомогательной информации коррекцию части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные наиболее высокому применимому уровню и любым уровням между наиболее высоким применимым уровнем и соответствующим уровнем. Базовое воссозданное представление звука может быть получено из компонентов, присвоенных наиболее высокому применимому уровню и любым уровням ниже наиболее высокого применимого уровня, с использованием базовой вспомогательной информации и скорректированной части дополнительной базовой вспомогательной информации, полученной из частей дополнительной базовой вспомогательной информации, соответствующих уровням вплоть до наиболее высокого применимого уровня. Способ может дополнительно содержать определение второго индекса уровня, который либо равен первому индексу уровня, либо указывает опущение улучшающей вспомогательной информации во время декодирования.

Сконфигурированный таким образом предложенный способ гарантирует, что дополнительная базовая вспомогательная информация, которая в конечном счете используется для декодирования базового сжатого представления звука, не включает в себя избыточные элементы, тем самым реализуя более эффективное фактическое декодирование базового сжатого представления звука.

Варианты осуществления этого аспекта могут относиться к вариантам осуществления предыдущего аспекта.

В соответствии с другим аспектом описан кодер для многоуровневого кодирования сжатого представления звука или звукового поля. Сжатое представление звука может включать в себя базовое сжатое представление звука, которое включает в себя множество компонентов. Множество компонентов могут являться



взаимодополняющими компонентами. Сжатое представление звука может дополнительно включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля. Сжатое представление звука может, кроме того, включать в себя улучшающую вспомогательную информацию, включающую в себя параметры для улучшения (например, расширения) базового воссозданного представления звука. Кодер может включать в себя процессор, выполненный с возможностью выполнять некоторые или все этапы способов в соответствии с первым упомянутым выше аспектом и вторым упомянутым выше аспектом.

В соответствии с другим аспектом описан декодер для декодирования сжатого представления звука или звукового поля. Сжатое представление звука может быть закодировано во множестве иерархических уровней. Множество иерархических уровней может включать в себя базовый уровень и один или несколько иерархических возрастающих уровней. Множество уровней может иметь присвоенные им компоненты базового сжатого представления звука или звукового поля. Другими словами, множество уровней может включать в себя компоненты базовой сжатой вспомогательной информации. Компоненты могут быть присвоены соответствующим уровням в соответствующих группах компонентов. Множество компонентов могут являться взаимодополняющими компонентами. Базовый уровень может включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука. Каждый уровень может включать в себя часть улучшающей вспомогательной информации, включающей в себя параметры для улучшения (например, расширения) базового воссозданного представления звука, доступные из данных, включенных в соответствующий уровень и любые уровни ниже соответствующего уровня. Декодер может включать в себя процессор, выполненный с возможностью выполнять некоторые или все этапы способов в соответствии с третьим упомянутым выше аспектом и четвертым упомянутым выше аспектом.

В соответствии с другими аспектами способы, устройства и системы направлены на декодирование сжатого представления Higher

Order Ambisonics (HOA) звука или звукового поля (пространственного звука или звукового поля высшего порядка). Устройство может иметь приемник, выполненный с возможностью принимать, или способ может принимать битовый поток, содержащий сжатое представление HOA, соответствующее множеству иерархических уровней, которые включают в себя базовый уровень и один или более иерархических улучшающих уровней. Множество уровней имеют присвоенные им компоненты базового сжатого представления звука или звукового поля, компоненты присвоены соответствующим уровням в соответствующих группах компонентов. Устройство может иметь декодер, выполненный с возможностью декодировать, или способ может декодировать сжатое представление HOA на основе базовой вспомогательной информации, которая связана с базовым уровнем, и на основе улучшающей вспомогательной информации, которая связана с одним или более иерархическими улучшающими уровнями. Базовая вспомогательная информация может включать в себя базовую независимую вспомогательную информацию, относящуюся к первым индивидуальным монауральным сигналам, которые будут декодироваться независимо от других монауральных сигналов. Каждый из одного или более иерархических улучшающих уровней может включать в себя часть улучшающей вспомогательной информации, включающей в себя параметры для улучшения базового воссозданного представления звука, доступные из данных, включенных в соответствующие уровни и любые уровни ниже соответствующего уровня.

Базовая независимая вспомогательная информация может указывать, что первые индивидуальные монауральные сигналы представляют направленный сигнал с направлением падения. Базовая вспомогательная информация может дополнительно включать в себя базовую зависимую вспомогательную информацию, относящуюся ко вторым индивидуальным монауральным сигналам, которые будут декодироваться зависимо от других монауральных сигналов. Базовая зависимая вспомогательная информация может включать в себя основанные на векторах сигналы, которые распределены по направлениям в звуковом поле, причем распределение по направлениям определено посредством вектора. Компоненты вектора

установлены равными нулю и не являются частью сжатого векторного представления.

Компоненты базового сжатого представления звука могут соответствовать монауральным сигналам, которые представляют либо преобладающие звуковые сигналы, либо последовательности коэффициентов представления НОА. Битовый поток включает в себя полезные нагрузки данных, соответственно связанные с множеством иерархических уровней. Улучшающая вспомогательная информация может включать в себя параметры, относящиеся по меньшей мере к одному из перечисленного: пространственное предсказание, синтез направленных подполосных сигналов и параметрическое дублирование звукового окружения. Улучшающая вспомогательная информация может включать в себя информацию, которая делает возможным предсказание недостающих частей звука или звукового поля на основе направленных сигналов. Может быть дополнительно определено для каждого уровня, был ли соответствующий уровень принят корректно, и индекс уровня, который находится непосредственно ниже наиболее низкого уровня, который не был принят корректно.

В соответствии с другим аспектом описана программа. Программа может быть адаптирована для исполнения на процессоре и для выполнения некоторых или всех этапов способа, изложенных в настоящем документе, при ее исполнении на вычислительном устройстве.

В соответствии с еще одним аспектом описан запоминающий носитель. Запоминающий носитель может содержать программу, адаптированную для исполнения на процессоре и для выполнения некоторых или всех этапов способа, изложенных в настоящем документе, при ее исполнении на вычислительном устройстве.

Утверждения, сделанные в отношении любого из упомянутых выше аспектов или их вариантов осуществления, также относятся к соответствующим другим аспектам или их вариантам осуществления, как поймет специалист в области техники. Повторение этих утверждений для каждого аспекта или варианта осуществления было опущено для краткости.

Способы и устройства, включающие в себя предпочтительные

варианты осуществления, изложенные в настоящем документе, могут использоваться автономно или в сочетании с другими способами и системами, раскрытыми в этом документе. Кроме того, все аспекты способов и устройств, изложенные в настоящем документе, могут быть произвольным образом объединены. В частности, признаки пунктов формулы изобретения могут быть объединены друг с другом произвольным образом.

Этапы способов и признаки устройств могут являться взаимозаменяемыми различным образом. В частности, подробности раскрытого способа могут быть реализованы как устройство, выполненное с возможностью исполнять некоторые или все этапы способа, и наоборот, как поймет специалист в области техники.

#### КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Изобретение разъяснено ниже иллюстративным образом со ссылкой на прилагаемые чертежи.

Фиг. 1 - блок-схема последовательности этапов, иллюстрирующая пример способа многоуровневого кодирования в соответствии с вариантами осуществления раскрытия;

Фиг. 2 - блок-схема, схематично иллюстрирующая пример стадии кодера в соответствии с вариантами осуществления раскрытия;

Фиг. 3 - блок-схема последовательности этапов, иллюстрирующая пример способа декодирования сжатого представления звука или звукового поля, который был закодирован в множестве иерархических уровней, в соответствии с вариантами осуществления раскрытия;

Фиг. 4А и фиг. 4В - блок-схемы, схематично иллюстрирующие примеры стадии декодера, в соответствии с вариантами осуществления раскрытия;

Фиг. 5 - блок-схема, схематично иллюстрирующая пример аппаратной реализации кодера в соответствии с вариантами осуществления раскрытия; и

Фиг. 6 - блок-схема, схематично иллюстрирующая пример аппаратной реализации декодера в соответствии с вариантами осуществления раскрытия.

#### ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

Сначала будет описано сжатое представление звука (или звукового поля) (далее для краткости называемое сжатым представлением звука), к которому применимы способы и кодеры/декодеры в соответствии с настоящим раскрытием. В целом полное сжатое представление звука (или звукового поля) (далее для краткости называемое полным сжатым представлением звука) может содержать три следующих компонента (например, состоять из них): базовое сжатое представление звука (или звукового поля) (далее для краткости называемое базовым сжатым представлением звука), базовую вспомогательную информацию и улучшающую вспомогательную информацию.

Само базовое сжатое представление звука содержит несколько компонентов (например, состоит из них) (например, взаимодополняющих компонентов). Базовое сжатое представление звука может принимать во внимание определенно наибольший процент полного сжатого представления звука. Базовое сжатое представление звука может состоять из монауральных транспортных сигналов, представляющих либо преобладающие звуковые сигналы, либо последовательности коэффициентов первоначального представления НОА.

Базовая вспомогательная информация нужна для декодирования базового сжатого представления звука и, как предполагается, имеет намного меньший размер по сравнению с базовым сжатым представлением звука. Это может быть сделано вплоть до ее наибольшей части несвязных частей, каждая из которых определяет восстановление только одного конкретного компонента базового сжатого представления звука. Базовая вспомогательная информация может содержать первую часть, которая может быть известна как независимая базовая вспомогательная информация, и вторую часть, которая может быть известна как дополнительная базовая вспомогательная информация.

И первая, и вторая части, независимая базовая вспомогательная информация и дополнительная базовая вспомогательная информация, могут определять восстановление конкретных компонентов базового сжатого представления звука. Вторая часть является факультативной и может быть опущена. В

этом случае можно сказать, что сжатое представление звука содержит первую часть (например, базовую вспомогательную информацию).

Первая часть (например, базовая вспомогательная информация) может содержать вспомогательную информацию, описывающую индивидуальные (взаимодополняющие) компоненты базового сжатого представления звука, независимо от других (взаимодополняющих) компонентов. В частности, первая часть (например, базовая вспомогательная информация) может определять декодирование одного или более из множества компонентов индивидуально, независимо от других компонентов. Таким образом, первая часть может упоминаться как независимая базовая вспомогательная информация.

Вторая (факультативная) часть может содержать вспомогательную информацию, также известную как дополнительная базовая вспомогательная информация, может описывать индивидуальные (взаимодополняющие) компоненты базового сжатого представления звука в зависимости от других (взаимодополняющих) компонентов. Эта вторая часть может также упоминаться как зависимая базовая вспомогательная информация. В частности, зависимость может иметь следующие свойства:

- Зависимая базовая вспомогательная информация для каждого индивидуального (взаимодополняющего) компонента базового сжатого представления звука может достигать своей наибольшей степени, когда другие определенные (взаимодополняющие) компоненты не содержатся в базовом сжатом представлении звука.

- В случае, если дополнительные определенные (взаимодополняющие) компоненты добавлены к базовому сжатому представлению звука, зависимая базовая вспомогательная информация для рассматриваемого индивидуального (взаимодополняющего) компонента может стать подмножеством первоначальной зависимой базовой вспомогательной информации, тем самым сокращая ее размер.

Улучшающая вспомогательная информация также является факультативной. Она может использоваться для улучшения или расширения (например, параметрического улучшения или расширения)

базового сжатого представления звука. Ее размер, как может также предполагаться, намного меньше, чем у базового сжатого представления звука.

Таким образом, в вариантах осуществления сжатое представление звука может содержать базовое сжатое представление звука, содержащее множество компонентов, базовую вспомогательную информацию для декодирования (например, восстановления) базового сжатого представления звука до базового воссозданного представления звука или звукового поля и улучшающую вспомогательную информацию, включающую в себя параметры для улучшения или расширения (например, параметрического улучшения или расширения) базового воссозданного представления звука. Сжатое представление звука может также содержать дополнительную базовую вспомогательную информацию для декодирования (например, восстановления) базового сжатого представления звука до базового воссозданного представления звука, которая может включать в себя информацию, которая определяет декодирование одного или более из множества компонентов в зависимости от соответствующих других компонентов.

Один пример такого типа полного сжатого представления звука задан посредством сжатого представления Higher Order Ambisonics (НОА) звукового поля (пространственного звукового поля высшего порядка), как определено посредством предварительной версии аудио стандарта MPEG-H 3D (ссылка 1), глава 12 и приложение С.5. Таким образом, сжатое представление звука может соответствовать сжатому представлению НОА звука (или звукового поля).

Для этого примера базовое сжатое представление звукового поля (базовое сжатое представление звука) может содержать несколько компонентов (например, может быть идентифицировано с их помощью). Компоненты могут представлять собой монауральные сигналы (например, соответствовать им). Монауральные сигналы могут представлять собой квантованные монауральные сигналы. Монауральные сигналы могут представлять либо преобладающие звуковые сигналы, либо последовательности коэффициентов окружающего компонента НОА звукового поля.

Базовая вспомогательная информация может описывать, среди

прочего, для каждого из этих монауральных сигналов, каким образом он вносит пространственный вклад в звуковое поле. Например, базовая вспомогательная информация может определять преобладающий звуковой сигнал как чисто направленный сигнал, означающий общую плоскую волну с некоторым направлением падения. В качестве альтернативы базовая вспомогательная информация может определять монауральный сигнал как последовательность коэффициентов первоначального представления НОА, имеющую некоторый индекс. Базовая вспомогательная информация также может быть разделена на первую часть и вторую часть, как указано выше.

Первая часть является вспомогательной информацией (например, независимой базовой вспомогательной информацией), относящейся к конкретным индивидуальным монауральным сигналам. Эта независимая базовая вспомогательная информация независима от существования других монауральных сигналов. Такая вспомогательная информация может, например, определять монауральный сигнал для представления направленного сигнала (например, означающего общую плоскую волну) с некоторым направлением падения. В качестве альтернативы монауральный сигнал может быть определен как последовательность коэффициентов первоначального представления НОА, имеющую некоторый индекс. Первая часть может упоминаться как независимая базовая вспомогательная информация. В целом первая часть (например, базовая вспомогательная информация) может определять декодирование одного или более из множества монауральных сигналов индивидуально, независимо от других монауральных сигналов.

Вторая часть является вспомогательной информацией (например, дополнительной базовой вспомогательной информацией), относящейся к конкретным индивидуальным монауральным сигналам. Эта вспомогательная информация зависит от существования других монауральных сигналов. Такая вспомогательная информация может быть использована, например, если монауральные сигналы определены как основанные на векторах сигналы (см., например, ссылку 1, раздел 12.4.2.4.4). Эти сигналы распределены по направлениям в звуковом поле, причем распределение по



направлениям может быть определено посредством вектора. В некотором режиме (см., например, `CodedVVecLength=1`) отдельные компоненты этого вектора неявно установлены равными нулю и не являются частью сжатого векторного представления. Этими компонентами являются компоненты с индексами, равными индексам последовательностей коэффициентов первоначального представления НОА и части базового сжатого представления звука. Это означает, что если индивидуальные компоненты вектора закодированы, их общее количество может зависеть от базового сжатого представления звука. В частности, общее количество может зависеть от того, какие последовательности коэффициентов содержит первоначальное представление НОА.

Если последовательности коэффициентов первоначального представления НОА не содержатся в базовом сжатом представлении звука, зависящая базовая вспомогательная информация для каждого основанного на векторе сигнала состоит из всех векторных компонентов и имеет свой наибольший размер. В случае, если последовательности коэффициентов первоначального представления НОА с некоторыми индексами добавляются к базовому сжатому представлению звука, векторные компоненты с этими индексами удаляются из вспомогательной информации для каждого основанного на векторе сигнала, тем самым сокращая размер зависимой базовой вспомогательной информации для основанных на векторах сигналов.

Улучшающая вспомогательная информация (например, улучшающая вспомогательная информация) может содержать параметры, относящиеся к (широкополосному) пространственному предсказанию (см. ссылку 1, раздел 12.4.2.4.3), и/или параметры, относящиеся к синтезу направленных подполосных сигналов и параметрическому дублированию звукового окружения.

Параметры, относящиеся к (широкополосному) пространственному предсказанию может использоваться для (линейного) предсказания недостающих частей звукового поля из направленных сигналов.

Синтез направленных подполосных сигналов и параметрическое дублирование звукового окружения являются инструментами сжатия, которые были недавно введены в аудио стандарт MPEG-H 3D с

помощью поправки [см. ссылку 2, раздел 1]. Эти два инструмента позволяют зависимому от частоты параметрическому предсказанию дополнительных монауральных сигналов быть пространственно распределенным, чтобы дополнять пространственно неполное или несовершенным образом сжатое представление НОА. Предсказание может быть основано на последовательностях коэффициентов базового сжатого представления звука.

Важно отметить, что упомянутый выше взаимодополняющий вклад в звуковое поле представлен в сжатом представлении НОА не посредством дополнительных квантованных сигналов, а посредством дополнительной вспомогательной информации сравнительно намного меньшего размера. Следовательно, два упомянутых инструмента кодирования особенно подходят для сжатия представлений НОА на низких скоростях передачи данных.

Второй пример сжатого представления одного или более монауральных сигналов с упомянутой выше структурой может содержать закодированную спектральную информацию для несвязных частотных полос вплоть до некоторой верхней частоты, что может рассматриваться как базовое сжатое представление; базовую вспомогательную информацию, определяющую закодированную спектральную информацию (например, посредством количества и ширины закодированных частотных полос); и улучшающую вспомогательную информацию содержащую параметры копирования спектральной полосы (SBR) (например, состоящую из них), которые описывают, как параметрически воссоздать из базового сжатого представления спектральную информацию для полос более высокой частоты, которые не рассматриваются в базовом сжатом представлении.

Настоящее раскрытие предлагает способ многоуровневого кодирования полного сжатого представления звука (или звукового поля), имеющего упомянутую выше структуру.

Сжатие может быть основано на кадрах в том смысле, что оно обеспечивает сжатые представления (в форме пакетов данных, или эквивалентно полезной нагрузки кадров) для последовательных временных интервалов. Временные интервалы могут иметь равные или разные размеры. Эти пакеты данных, как может предполагаться,

содержат флаг корректности, значение, указывающее их размер, а также фактические данные сжатого представления. Далее без намеренного ограничения будет предполагаться, что сжатие является основанным на кадрах. Кроме того, если не указано иначе, и без намеренного ограничения будет сделан фокус на обработке одного кадра, и поэтому индекс кадра будет опущен.

Каждая полезная нагрузка кадра рассматриваемого полного сжатого представления звука (или звукового поля), как предполагается, содержит  $J$  пакетов данных (или полезных нагрузок кадра), каждый для одного компонента базового сжатого представления звука, которые обозначены как  $BSRC_j$ ,  $j = 1, \dots, J$ . Кроме того, предполагается, что пакет содержит *независимую* базовую вспомогательную информацию (базовую вспомогательную информацию), обозначенную как  $BSI_I$ , определяющую отдельные компоненты  $BSRC_j$  базового сжатого представления звука, независимо от других компонентов. Факультативно может дополнительно предполагаться, что пакет содержит *зависимую* базовую вспомогательную информацию (дополнительную базовую вспомогательную информацию), обозначенную как  $BSI_D$ , определяющую отдельные компоненты  $BSRC_j$  базового сжатого представления звука в зависимости от других компонентов.

Информация, содержащаяся в двух пакетах данных  $BSI_I$  и  $BSI_D$ , может быть факультативно сгруппирована в единственный пакет данных  $BSI$  базовой вспомогательной информации. Можно сказать, что единственный пакет данных  $BSI$  содержит, среди прочего,  $J$  частей, каждая из которых определяет один отдельный компонент  $BSRC_j$  базового сжатого представления звука. Можно сказать, что каждая из этих частей, в свою очередь, содержит часть независимой вспомогательной информации и факультативно часть зависимой вспомогательной информации.

В конечном счете, она может включать в себя полезную нагрузку улучшающей вспомогательной информации (улучшающей вспомогательной информации), обозначенную как  $ESI$ , с описанием того, как улучшить или расширить воссозданный звук (или звуковое

поле) на основе полного базового сжатого представления звука.

Предлагаемое решение для многоуровневого кодирования направлено на этапы, требующиеся для обеспечения возможности как для части сжатия, включающей в себя упаковку пакетов данных для передачи, а также для части приема и восстановления. Каждая часть будет подробно описана далее.

Сначала будут описаны сжатие и упаковка (например, для передачи). В частности, будут описаны компоненты и элементы полного сжатого представления звука (или звукового поля) в случае многоуровневого кодирования.

Фиг. 1 схематично иллюстрирует блок-схему последовательности этапов примера способа сжатия и упаковки (например, способа кодирования или способа многоуровневого кодирования сжатого представления звука или звукового поля). Присвоение (например, распределение) индивидуальных полезных нагрузок базовому уровню и  $(M-1)$  улучшающим уровням может быть достигнуто посредством упаковщика транспортных уровней. Фиг. 2 схематично иллюстрирует блок-схему примера присвоения/распределения индивидуальных полезных нагрузок.

Как указано выше, полное сжатое представление 2100 звука может относиться, например, к сжатому представлению НОА, содержащему базовое сжатое представление звука. Полное сжатое представление 2100 звука может содержать множество компонентов (например, монауральные сигналы) 2110-1, ... 2110- $J$ , независимую базовую вспомогательную информацию (базовую вспомогательную информацию) 2120, факультативную улучшающую вспомогательную информацию (улучшающую вспомогательную информацию) 2140 и факультативную зависимую базовую вспомогательную информацию (дополнительную базовую вспомогательную информацию) 2130. Базовая вспомогательная информация 2120 может являться информацией для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля. Базовая вспомогательная информация 2120 может включать в себя информацию, которая определяет декодирование одного или более компонентов (например, монауральных сигналов)

индивидуально, независимо от других компонентов. Улучшающая вспомогательная информация 2140 может включать в себя параметры для улучшения (например, расширения) базового воссозданного представления звука. Дополнительная базовая вспомогательная информация 2130 может являться (дополнительной) информацией для декодирования базового сжатого представления звука в базовое воссозданное представление звука и может включать в себя информацию, которая определяет декодирование одного или более из множества компонентов в зависимости от соответствующих других компонентов.

Фиг. 2 иллюстрирует основополагающее допущение, в котором существует множество иерархических уровней, включающих в себя один базовый уровень (основной уровень) и один или более (иерархических) улучшающих уровней. Например, может иметься всего  $M$  уровней, т.е. один базовый уровень и  $M-1$  улучшающих уровней. Множество иерархических уровней имеет последовательно увеличивающийся индекс уровня. Самое низкое значение индекса уровня (например, индекс 1 уровня) соответствует базовому уровню. Далее подразумевается, что уровни упорядочены от базового уровня, через улучшающие уровни, вплоть до полного наиболее высокого улучшающего уровня (т.е., полного наиболее высокого уровня).

Предложенный способ может быть выполнен на основе кадра (т.е., покадрово). В частности, сжатое представление 2100 звука может быть сжато для последовательных временных интервалов, например, временных интервалов равного размера. Каждый временной интервал может соответствовать кадру. Описанные ниже этапы могут быть выполнены для каждого последовательного временного интервала (например, кадра).

На этапе S1010 на фиг. 1 множество компонентов 2110 подразделяется на множество групп компонентов. Каждая из множества групп затем присваивается (например, добавляется или распределяется) соответствующему одному из множества иерархических уровней. При этом количество групп соответствует количеству уровней. Например, количество групп может быть равно

количеству уровней, чтобы имелась одна группа компонентов для каждого уровня. Как указано выше, множество уровней может включать в себя базовый уровень и один или более (например,  $M-1$ ) иерархических улучшающих уровней.

Другими словами, базовое сжатое представление звука подразделено на части, которые будут присвоены отдельным уровням. Без потери общности группировка может быть описана посредством  $M+1$  чисел  $J_m$ ,  $m=0, \dots, M$ , где  $J_0=1$  и  $J_M=J+1$ , в результате чего компоненты  $BSRC_j$  присваиваются  $m$ -ому уровню для  $J_{m-1} \leq j < J_m$ .

На этапе S1020 группы компонентов присваиваются своим соответствующим уровням. На этапе S1030 базовая вспомогательная информация 2120 добавляется (например, распределяется) к базовому уровню (т.е., к наиболее низкому из множества иерархических уровней).

Таким образом, вследствие ее небольшого размера предложено включать полную базовую вспомогательную информацию (базовую вспомогательную информацию и факультативную дополнительную базовую вспомогательную информацию) в базовый уровень, чтобы избежать ее ненужной фрагментации.

Если рассматриваемое сжатое представление звука содержит зависимую базовую вспомогательную информацию (дополнительную базовую вспомогательную информацию), способ дополнительно может содержать (не показано на фиг. 1) декомпозицию дополнительной базовой вспомогательной информации на множество частей 2130-1, ..., 2130- $M$  дополнительной базовой вспомогательной информации. Части дополнительной базовой вспомогательной информации затем могут быть добавлены (например, распределены) к базовому уровню. Другими словами, части дополнительной базовой вспомогательной информации могут быть включены в базовый уровень. Каждая часть дополнительной базовой вспомогательной информации может быть связана с соответствующим уровнем и может включать в себя информацию, которая определяет декодирование одного или более компонентов, присвоенных соответствующему уровню, в зависимости

от других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня.

Таким образом, в то время как независимая базовая вспомогательная информация  $BSI_I$  (базовая вспомогательная информация) 2120 оставляется без изменений для присвоения, зависимая базовая вспомогательная информация должна быть обработана специально для многоуровневого кодирования, чтобы позволить правильное декодирование на стороне приемника, с одной стороны, и сократить размер зависимой базовой вспомогательной информации для передачи, с другой стороны. Предложено выполнить декомпозицию зависимой базовой вспомогательной информации на  $M$  частей, обозначенных как  $BSI_{D,m}$ ,  $m = 1, \dots, M$ , где  $m$ -ая часть содержит зависимую базовую вспомогательную информацию для каждого из компонентов  $BSRC_j$ ,  $J_{m-1} \leq j < J_m$  базового сжатого представления звука, присвоенного  $m$ -ому уровню, в предположении, что факультативная зависимая базовая вспомогательная информация существует для рассматриваемого сжатого представления звука. В случае, если соответствующая зависимая вспомогательная информация не существует, для сжатого представления звука частей  $BSI_{D,m}$  может предполагаться пустой. Каждая часть зависимой базовой вспомогательной информации  $BSI_{D,m}$  может зависеть от всех компонентов  $BSRC_j$ ,  $1 \leq j < J_m$ , содержащихся на всех уровнях вплоть до  $m$ -ого (т.е., содержащихся на всех уровнях  $j = 1, \dots, m$ ).

Если пакет  $BSI_I$  независимой базовой вспомогательной информации имеет пренебрежительно небольшой размер, разумно удерживать его как целое и добавлять (присваивать) его к базовому уровню. Факультативно подобная декомпозиция, как для зависимой базовой вспомогательной информации, также может быть выполнена для независимой базовой вспомогательной информации, обеспечивая пакеты  $BSI_{I,m}$ ,  $m = 1, \dots, M$ . Это полезно для сокращения размера базового уровня посредством добавления (присвоения) частей независимой базовой вспомогательной информации к уровням

с соответствующими компонентами базового сжатого представления звука.

На этапе S1040 может быть определено множество частей 2140-1, ..., 2140- $M$  улучшающей вспомогательной информации. Каждая часть улучшающей вспомогательной информации может включать в себя параметры для улучшения (например, расширения) воссозданного представления звука, доступные из данных, включенных в соответствующий уровень и любые уровни ниже соответствующего уровня.

Причина выполнения этого этапа состоит в том, что в случае многоуровневого кодирования важно реализовать, чтобы улучшающая вспомогательная информация должна была вычисляться для каждого дополнительного уровня, поскольку предполагается улучшить предварительный восстановленный звук (или звуковое поле), что, однако, зависит от доступных уровней для восстановления. В частности, предварительный восстановленный звук (или звуковое поле) для данного наиболее высокого декодируемого уровня (наиболее высокого применимого уровня) зависит от компонентов, включенных в наиболее высокий декодируемый уровень и любые уровни ниже наиболее высокого декодируемого уровня. Следовательно, сжатие должно обеспечить  $M$  индивидуальных пакетов данных улучшающей вспомогательной информации (частей улучшающей вспомогательной информации), обозначенных как  $ESI_m$ ,  $m = 1, \dots, M$ , где улучшающая вспомогательная информация в  $m$ -ом пакете данных  $ESI_m$  вычисляется, чтобы улучшить представление звука (или звукового поля), полученное из всех данных, содержащихся на базовом уровне и улучшающих уровнях с индексами ниже  $m$  (например, всех данных, содержащихся на  $m$ -ом уровне и любых уровнях ниже  $m$ -ого уровня).

На этапе S1050 множество частей 2140-1, ..., 2140- $M$  улучшающей вспомогательной информации присваивается (например, добавлено или распределяется) множеству уровней. Каждая из множества частей улучшающей вспомогательной информации присваивается соответствующему одному из множества уровней. Например, каждый



из множества уровней включает в себя соответствующую часть улучшающей вспомогательной информации.

Присвоение базовой и/или улучшающей вспомогательной информации соответствующим уровням может быть указано в информации конфигурации, которая формируется посредством способа кодирования. Другими словами, соответствие между базовой и/или улучшающей вспомогательной информацией и соответствующими уровнями может быть указано в информации конфигурации. Кроме того, информация конфигурации может указывать для каждого уровня компоненты базового сжатого представления звука, которые присвоены (например, включены) этому уровню. Части дополнительной базовой вспомогательной информации, включенные в базовый уровень, все же могут соответствовать уровням, отличающимся от базового уровня.

Подводя итог, на стадии сжатия обеспечивается пакет данных кадра, обозначенный как **FRAME**, который имеет следующий состав:

$$\mathbf{FRAME} = [\mathbf{BSRC}_1 \dots \mathbf{BSRC}_j \mathbf{BSI}_1 \mathbf{BSI}_{D,1} \dots \mathbf{BSI}_{D,M} \mathbf{ESI}_1 \dots \mathbf{ESI}_M] \quad (1)$$

Кроме того, пакеты  $\mathbf{BSI}_1$  и  $\mathbf{BSI}_{D,m}$  для  $m = 1, \dots, M$  могут бы быть объединены в единственный пакет  $\mathbf{BSI}$ , в этом случае пакет данных кадра, обозначенный как **FRAME**, будет иметь следующий состав:

$$\mathbf{FRAME} = [\mathbf{BSRC}_1 \mathbf{BSRC}_2 \dots \mathbf{BSRC}_j \mathbf{BSI} \mathbf{ESI}_1 \mathbf{ESI}_2 \dots \mathbf{ESI}_M] \quad (2)$$

Порядок следования индивидуальных полезных нагрузок с пакетом данных кадра в общем случае может быть произвольным.

Индивидуальные пакеты данных затем могут быть сгруппированы в полезные нагрузки, которые определены как специальные пакеты данных, которые содержат флаг корректности, значение, указывающее их размер, а также фактические сжатые данные представления. Использование полезных нагрузок позволяет простое демультиплексирование на стороне приемника, предлагая преимущество возможности отбрасывать неактуальные полезные нагрузки без необходимости их анализа. Одна возможная группировка задана как

- присвоение (например, распределение) каждого  $\mathbf{BSRC}_j$  пакета  $j = 1, \dots, J$  индивидуальной полезной нагрузке, обозначенной как  $\overline{\mathbf{BP}}_j$ .

- присвоение (например, распределение)  $m$ -ого пакета данных улучшающей вспомогательной информации  $ESI_m$  и  $m$ -го пакета данных  $BSI_{D,m}$  зависимой вспомогательной информации одной улучшающей полезной нагрузке, обозначенной как  $\overline{EP}_m$ ,  $m = 1, \dots, M$ .

- присвоение пакета независимой базовой вспомогательной информации  $BSI_I$  отдельной полезной нагрузке вспомогательной информации, обозначенной как  $\overline{BSIP}$ .

Факультативно, если размер независимой базовой вспомогательной информации большой, каждый  $m$ -ый из ее компонентов,  $BSI_{I,m}$ ,  $m = 1, \dots, M$ , может быть присвоен (например, распределен) улучшающей полезной нагрузке  $\overline{EP}_m$ . В этом случае полезная нагрузка  $\overline{BSIP}$  вспомогательной информации является пустой и может быть проигнорирована.

Другая факультативная возможность состоит в том, чтобы присвоить все зависимые пакеты данных  $BSI_{D,m}$  базовой вспомогательной информации полезной нагрузке  $\overline{BSIP}$  вспомогательной информации, что является разумным, если размер зависимой базовой вспомогательной информации является небольшим.

В конечном счете может быть обеспечен пакет данных кадра, обозначенный как  $FRAME$ , имеющий следующий состав

$$FRAME = [\overline{BP}_1 \dots \overline{BP}_I \overline{BSIP} \overline{EP}_1 \dots \overline{EP}_M] \quad (3)$$

Порядок следования индивидуальных полезных нагрузок с пакетом данных кадра в общем случае может быть произвольным.

Способ может дополнительно содержать (не показано на фиг. 1) формирование для каждого из множества уровней пакета транспортного уровня (например, пакета 2200 базового уровня и  $M-1$  пакетов 2300-1, ..., 2300-( $M-1$ )) улучшающего уровня, включающих в себя данные соответствующего уровня (например, компоненты, базовую вспомогательную информацию и улучшающую вспомогательную информацию для базового уровня, или компоненты и улучшающую вспомогательную информацию для одного или более улучшающих уровней).

Пакеты транспортного уровня для разных уровней могут иметь разные приоритеты передачи. Таким образом, способ может дополнительно содержать (не показано на фиг. 1) формирование транспортного потока для передачи данных множества уровней, причем базовый уровень имеет наиболее высокий приоритет передачи, и иерархические улучшающие уровни имеют убывающие приоритеты передачи. При этом более высокий приоритет передачи может соответствовать большей степени защиты от ошибок, и наоборот.

Если этапы не требуют некоторых других этапов в качестве предварительных условий, упомянутые выше этапы могут выполняться в любом порядке, и предполагается, что иллюстративный порядок, показанный на фиг. 1, не имеет ограничительного характера.

Фиг. 3 иллюстрирует способ декодирования сжатого представления звука или звукового поля для декодирования или восстановления. Примеры соответствующей стадии приема и восстановления схематично проиллюстрированы на блок-схемах на фиг. 4А и фиг. 4В.

Как следует из предыдущего описания, сжатое представление звука может быть закодировано в множестве иерархических уровней. Множество уровней может иметь присвоенные им (например, может включать в себя) компоненты базового сжатого представления звука, компоненты присваиваются соответствующим уровням в соответствующих группах компонентов. Базовый уровень может включать в себя базовую вспомогательную информацию для декодирования базового сжатого представления звука. Каждый уровень может включать в себя одну из упомянутых выше частей улучшающей вспомогательной информации, включающей в себя параметры для улучшения базового воссозданного представления звука, доступных из данных, включенных в соответствующий уровень и любые уровни ниже соответствующего уровня.

Предложенный способ может быть выполнен на основе кадров (т.е., покадрово). В частности, восстановленное представление звука или звукового поля может быть сформировано для последовательных временных интервалов, например, временных интервалов равного размера. Временные интервалы могут являться,

например, кадрами. Описанные ниже этапы могут быть выполнены для каждого последовательных временных интервалов (например, кадров).

На этапе S3010 принимаются полезные нагрузки данных (например, пакеты транспортного уровня), соответствующие множеству уровней. Полезные нагрузки данных могут быть приняты как часть битового потока, который содержит сжатое представление НОА звука или звукового поля, представление соответствует множеству иерархических уровней. Иерархические уровни включают в себя базовый уровень и один или более иерархических улучшающих уровней. Множество уровней имеет присвоенные им компоненты базового сжатого представления звука или звукового поля. Компоненты присвоены соответствующим уровням в соответствующих группах компонентов.

Пакеты индивидуальных уровней могут быть мультиплексированы для обеспечения принятого пакета кадра полного сжатого представления звука. Принятый пакет кадра может быть обозначен как

$$[BSI_I \ BSI_{D,1} \ \dots \ BSI_{D,M} \ ESI_1 \ BSRC_1 \ \dots \ BSRC_{(J_1)-1} \ \dots \ ESI_M]$$

(4)

В альтернативном случае пакеты  $BSI_I$  и  $BSI_{D,m}$  для  $m = 1, \dots, M$  объединены в единственный пакет  $BSI$ , пакеты индивидуальных уровней могут быть мультиплексированы для обеспечения принятого пакета кадра полного сжатого представления звука, обозначенного как

$$[BSI \ ESI_1 \ BSRC_1 \ \dots \ BSRC_{(J_1)-1} \ \dots \ ESI_M \ BSRC_{J_{(M-1)}} \ \dots \ BSRC_J]$$
 (5)

В терминах полезных нагрузок принятый пакет кадра может быть задан как

$$FRAME = [\overline{BP}_1 \ \dots \ \overline{BP}_1 \ \overline{BSIP} \ \overline{EP}_1 \ \dots \ \overline{EP}_M]$$
 (6)

Принятый пакет кадра затем может быть передан на декомпрессор или декодер 4100. Если передача индивидуального уровня была безошибочной, флаг корректности по меньшей мере части содержащейся полезной нагрузки  $\overline{EP}_m$  улучшающей вспомогательной информации (например, соответствующей части

улучшающей вспомогательной информации) установлен равным "истинному". В случае ошибки вследствие передачи индивидуального уровня флаг корректности по меньшей мере в полезной нагрузке улучшающей вспомогательной информации на этом уровне установлен равным "ложному". Следовательно, корректность пакета уровня может быть определена на основе корректности содержащейся полезной нагрузки улучшающей вспомогательной информации (например, на основе ее флага корректности).

В декомпрессоре 4100 принятый пакет кадра может быть демультимплексирован. С этой целью может использоваться информация размера каждой полезной нагрузки, чтобы избежать ненужного анализа данных индивидуальных полезных нагрузок.

На этапе S3020 первый индекс уровня, указывающий наиболее высокий уровень (например, наиболее высокий применимый уровень или наиболее высокий декодируемый уровень), определяется из множества уровней для использования для декодирования базового сжатого представления звука в базовое воссозданное представление звука или звукового поля.

Кроме того, на этапе S3020 может быть выбрано значение (например, индекс уровня)  $N_B$  наиболее высокого уровня (наиболее высокого применимого уровня), который будет использоваться для восстановления базового представления звука. Наиболее высокий *улучшающий* уровень, который будет фактически использоваться для восстановления базового представления звука, задан как  $N_B - 1$ . Поскольку каждый уровень содержит точно одну полезную нагрузку улучшающей вспомогательной информации (часть улучшающей вспомогательной информации), можно определить на основе полезной нагрузки улучшающей вспомогательной информации, является ли корректным содержащий уровень (например, был корректно принят). Следовательно, выбор может быть достигнут с использованием всех полезных нагрузок улучшающей вспомогательной информации  $ESI_m$ ,  $m = 1, \dots, M$  (или, соответственно,  $\overline{EP}_m$ ,  $m = 1, \dots, M$ ).

На этапе S3030 получается базовое воссозданное представление звука. Базовое воссозданное представление звука

может быть получено из компонентов, присвоенных наиболее высокому применимому уровню, указанному первым индексом уровня, и любым уровням ниже этого наиболее высокого применимого уровня с использованием базовой вспомогательной информации (или в целом с использованием базовой вспомогательной информации).

Полезные нагрузки компонентов  $BSRC_1, \dots, BSRC_J$  базового сжатого представления звука могут быть обеспечены наряду с (всеми) полезными нагрузками базовой вспомогательной информации (например,  $BSI$  или  $BSI_1$  и  $BSI_{D,m}$ ,  $m = 1, \dots, M$ ) и значением  $N_B$  процессору 4200 восстановления базового представления. Процессор 4200 восстановления базового представления (проиллюстрированный на фиг. 4А и 4В) воссоздает базовое представление звука (или звукового поля) с использованием только тех компонентов базового сжатого представления звука, которые содержатся на наиболее низких  $N_B$  уровнях, которые представляют собой базовый уровень и  $N_B - 1$  улучшающих уровней (т.е., уровни вплоть до уровня, указанного первым индексом уровня). В качестве альтернативы процессору 4200 восстановления базового представления могут быть обеспечены только полезные нагрузки компонентов базового сжатого представления звука, содержащиеся на наиболее низких  $N_B$  уровнях вместе с соответствующими полезными нагрузками базовой вспомогательной информации.

Требуемая информация о том, какие компоненты базового сжатого представления звука (или звукового поля) содержатся на индивидуальных уровнях, предполагается известной декомпрессору 4100 из пакета данных с информацией конфигурации, которая предполагается отправленной и принятой перед пакетами данных кадра.

Чтобы обеспечить пакеты данных  $BSI_{D,m}$ ,  $m = 1, \dots, N_B$  зависимой вспомогательной информации и пакет данных  $ESI_{N_E}$  улучшающей вспомогательной информации, все улучшающие полезные нагрузки могут быть введены частичный анализатор 4400 (см. фиг. 4В) декомпрессора 4100 вместе со значением  $N_E$  и значением  $N_B$ .

Анализатор может отбросить все полезные нагрузки и пакеты данных, которые не будут использоваться для фактического восстановления. Если значение  $N_E$  равно нулю, то может предполагаться, что все пакеты данных улучшающей вспомогательной информации являются пустыми.

Если базовый уровень включает в себя по меньшей мере одну зависимую полезную нагрузку базовой вспомогательной информации (часть дополнительной базовой вспомогательной информации), соответствующей соответствующему уровню, декодирование каждой индивидуальной полезной нагрузки зависимой базовой вспомогательной информации (например,  $BSI_{D,m}$ ,  $m = 1, \dots, N_B$  (часть дополнительной базовой вспомогательной информации)) может включать в себя (i) декодирование части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные ее соответствующему уровню и любым уровням ниже соответствующего уровня (предварительное декодирование), и (ii) коррекцию части дополнительной базовой вспомогательной информации посредством ссылки на компоненты, присвоенные наиболее высокому применимому уровню и любым уровням между наиболее высоким применимым уровнем и соответствующим уровнем (коррекция). При этом дополнительная базовая вспомогательная информация, соответствующая соответствующему уровню, включает в себя информацию, которая определяет декодирование одного или более компонентов среди компонентов, присвоенных соответствующему уровню, в зависимости от других компонентов, присвоенных соответствующему уровню и любым уровням ниже соответствующего уровня.

Затем базовое воссозданное представление звука может быть получено (например, сформировано) из компонентов, присвоенных наиболее высокому применимому уровню и любым уровням ниже наиболее высокого применимого уровня, с использованием базовой вспомогательной информации и скорректированных частей дополнительной базовой вспомогательной информации, полученных из частей дополнительной базовой вспомогательной информации, соответствующей уровням вплоть до наиболее высокого применимого

уровня.

В частности, предварительное декодирование каждой полезной нагрузки  $BSI_{D,m}$ ,  $m = 1, \dots, N_B$ , может включать в себя использование ее зависимость от первых  $J_m - 1$  компонентов  $BSRC_1, \dots, BSRC_{(J_m) - 1}$  базового сжатого представления звука, содержащихся на первых  $m$  уровнях, что предполагалось на стадии кодирования.

Последовательная коррекция каждой полезной нагрузки  $BSI_{D,m}$ ,  $m = 1, \dots, N_B$ , может включать в себя принятие во внимание, что базовый компонент звука наконец воссоздан из первых  $J_{N_B} - 1$  компонентов  $BSRC_1, \dots, BSRC_{(J_{N_B}) - 1}$  базового сжатого представления звука, содержащихся на первых  $N_B > m$  уровнях, что является большим количеством компонентов, чем предполагалось для предварительного декодирования. Следовательно, коррекция может быть достигнута посредством отбрасывания неадекватной информации, что возможно вследствие первоначально принятого свойства зависимой базовой вспомогательной информации, состоящего в том, что если некоторые взаимодополняющие компоненты добавляются к базовому сжатому представлению звука, зависимая базовая вспомогательная информация для каждого индивидуального (взаимодополняющего) компонента становится подмножеством первоначальной.

На этапе S3040 может быть определен второй индекс уровня. Второй индекс уровня может указывать часть (части) улучшающей вспомогательной информации, которая должна использоваться для улучшения (например, расширения) базового воссозданного представления звука.

В дополнение к первому индексу уровня может быть определен индекс  $N_E$  (второй индекс уровня) полезной нагрузки улучшающей вспомогательной информации (части второй улучшающей информации) для использования для восстановления. Второй индекс  $N_E$  уровня может всегда либо быть равен первому индексу  $N_B$  уровня, либо быть равен нулю. Улучшение может быть достигнуто либо всегда в



соответствии с базовым представлением звука, полученным из наиболее высокого применимого уровня, либо никогда.

На этапе S3050 воссозданное представление звука или звукового поля получается (например, формируется) из базового воссозданного представления звука со ссылкой на второй индекс уровня.

Таким образом, воссозданное представление звука получается посредством (параметрического) улучшения или расширения базового воссозданного представления звука, например, посредством использования улучшающей вспомогательной информации (части улучшающей вспомогательной информации), указанной вторым индексом уровня. Как указано далее, второй индекс уровня может указывать на то, чтобы вообще не использовать какую-либо улучшающую вспомогательную информацию на данной стадии. Тогда воссозданное представление звука будет соответствовать базовому воссозданному представлению звука.

С этой целью воссозданное базовое представление звука вместе со всеми полезными нагрузками  $ESI_1, \dots, ESI_M$  улучшающей вспомогательной информации, полезными нагрузками базовой вспомогательной информации (например,  $BSI$  или  $BSI_I$ , и  $BSI_{D,m}$ ,  $m = 1, \dots, M$ ), и значением  $N_E$  обеспечиваются процессору 4300 восстановления расширенного представления (проиллюстрированному на фиг. 4А и 4В), который вычисляет окончательное расширенное представление 2100' звука (или звукового поля) с использованием только полезной нагрузки  $ESI_{N_E}$  улучшающей вспомогательной информации, и отбрасывая все другие полезные нагрузки улучшающей вспомогательной информации. В качестве альтернативы процессору 4300 восстановления улучшающего представления может быть обеспечена только полезная нагрузка  $ESI_{N_E}$  улучшающей вспомогательной информации вместо всех полезных нагрузок улучшающей вспомогательной информации. Если значение  $N_E$  равно нулю, все полезные нагрузки улучшающей вспомогательной информации отбрасываются (или в качестве альтернативы полезная нагрузка улучшающей вспомогательной информации не

обеспечивается), и воссозданное финальное расширенное представление 2100' звука равно воссозданному основному представлению звука. Полезная нагрузка  $ESI_{NE}$  улучшающей вспомогательной информации может быть получена посредством частичного анализатора 4400.

Фиг. 3 также в целом иллюстрирует декодирование сжатого представления НОА на основе базовой вспомогательной информации, которая связана с базовым уровнем, и на основе улучшающей вспомогательной информации, которая связана с одним или более иерархическими улучшающими уровнями.

Если этапы не требуют некоторых других этапов в качестве предварительных условий, упомянутые выше этапы могут выполняться в любом порядке, и предполагается, что иллюстративный порядок, показанный на фиг. 3, не имеет ограничительного характера.

Далее будут описаны подробности выбора уровней для восстановления (выбор первого и второго индексов уровней) на этапах S3020 и S3040.

Определение первого индекса уровня может включать в себя определение для каждого уровня, был ли соответствующий уровень принят корректно. Определение первого индекса уровня может дополнительно включать в себя определение первого индекса уровня как индекса того уровня, который находится непосредственно ниже наиболее низкого уровня, который не был корректно принят. Был ли уровень принят корректно, может быть определено посредством оценки, была ли корректно принята полезная нагрузка улучшающей вспомогательной информации этого уровня. Это, в свою очередь, может быть сделано посредством оценки флагов корректности в полезных нагрузках улучшающей вспомогательной информации.

Определение второго индекса уровня в общем случае может включать в себя либо определение второго индекса уровня как равного первому индексу уровня, либо определение значения индекса в качестве второго индекса уровня (например, значение 0 индекса), которое указывает, что не следует использовать какую-либо улучшающую вспомогательную информацию при получении воссозданного представления звука.

В случае, если все пакеты данных кадра могут быть восстановлены независимо друг от друга, и номер  $N_B$  наиболее высокого уровня (наиболее высокого применимого уровня) для фактического использования для восстановления базового представления звука, и индекс  $N_E$  полезной нагрузки улучшающей вспомогательной информации для использования для восстановления могут быть установлены равными наибольшему номеру  $L$  корректной полезной нагрузки улучшающей вспомогательной информации, который сам может быть определен посредством оценки флагов корректности в полезных нагрузках улучшающей вспомогательной информации. Используя знание размера каждой полезной нагрузки улучшающей вспомогательной информации, можно избежать сложного анализа фактических данных полезных нагрузок для определения их корректности.

Таким образом, второй индекс уровня может быть определен как равный первому индексу уровня, если сжатые представления звука для последовательных временных интервалов могут быть декодированы независимо. В этом случае воссозданное базовое представление звука может быть расширено на основе полезной нагрузки улучшающей вспомогательной информации наиболее высокого применимого уровня.

В случае, если используется это дифференциальное восстановление с межкадровыми зависимостями, в дополнение следует рассматривать решение от предыдущего кадра. Следует отметить, что с дифференциальным восстановлением обычно независимые пакеты данных кадра передаются с регулярными временными интервалами, чтобы позволить начинать восстановление с тех моментов времени, когда определение значений  $N_B$  и  $N_E$  становятся независимыми от кадров, и она выполняется, как описано выше.

Для подробного разъяснения предложенного зависимого от кадров решения, самое большой номер (например, индекс уровня) корректной полезной нагрузки улучшающей вспомогательной информации для  $k$ -ого кадра обозначен как  $L(k)$ , номер наиболее

высокого уровня (например, индекс уровня) для выбора и использования для восстановления базового представления звука обозначен как  $N_B(k)$ , и номер (например, индекс уровня) полезной нагрузки улучшающей вспомогательной информации для использования для восстановления обозначен как  $N_E(k)$ .

Используя эти обозначения, номер наиболее высокого уровня для использования для восстановления базового представления звука  $N_B(k)$  может быть вычислен в соответствии с

$$N_B(k) = \min(N_B(k-1), L(k)). \quad (7)$$

Посредством выбора  $N_B(k)$  не больше, чем  $N_B(k-1)$  и  $L(k)$ , обеспечивается, что вся информация, требуемая для дифференциального восстановления базового представления звука, является доступной.

Таким образом, если сжатые представления звука для последовательных временных интервалов (например, кадров) не могут быть декодированы независимо друг от друга, определение первого индекса уровня может содержать определение для каждого уровня, был ли соответствующий уровень принят корректно, и определение первого индекса уровня для данного временного интервала как меньшего индекса из первого индекса уровня временного интервала, предшествующего данному временному интервалу, и индекса уровня, который находится непосредственно ниже наиболее низкого уровня, который не был корректно принят.

Номер  $N_E(k)$  полезной нагрузки улучшающей вспомогательной информации для использования для восстановления может быть определен в соответствии с

$$N_E(k) = \begin{cases} N_B(k) & \text{if } N_B(k) = N_B(k-1) \\ 0 & \text{else} \end{cases}. \quad (8)$$

При этом выбор 0 для  $N_E(k)$  указывает, что воссозданное базовое представление звука не должно улучшаться или расширяться с использованием улучшающей вспомогательной информации.

Это означает, в частности, что при условии, что номер  $N_B(k)$  наиболее высокого уровня для использования для восстановления

базового представления звука не изменяется, выбирается тот же самый соответствующий номер улучшающего уровня. Однако в случае изменения  $N_B(k)$  улучшение запрещается посредством установки  $N_E(k)$  равным нулю. Вследствие предполагаемого дифференциального восстановления улучшающей вспомогательной информации ее изменение в соответствии с  $N_B(k)$  невозможно, поскольку это потребовало бы восстановления соответствующего уровня улучшающей вспомогательной информации в предыдущем кадре, которая, как предполагается, не была выполнена.

Таким образом, если сжатые представления звука для последовательных временных интервалов (например, кадров) не могут быть декодированы независимо друг от друга, определение второго индекса уровня может содержать определение, равен ли первый индекс уровня для данного временного интервала первому индексу уровня для предыдущего временного интервала. Если первый индекс уровня для данного временного интервала равен первому индексу уровня для предыдущего временного интервала, второй индекс уровня для данного временного интервала может быть определен (например, выбран) как равный первому индексу уровня для данного временного интервала. С другой стороны, если первый индекс уровня для данного временного интервала не равен первому индексу уровня для предыдущего временного интервала, значение индекса может быть определено (например, выбрано) как второй индекс уровня, который указывает, что не следует использовать какую-либо улучшающую вспомогательную информацию при получении воссозданного представления звука.

В качестве альтернативы, если при восстановлении все полезные нагрузки улучшающей вспомогательной информации с номером вплоть до  $N_E(k)$  восстановлены параллельно, правило выбора в уравнении (4) может быть заменено:

$$N_E(k) = N_B(k). \quad (9)$$

Наконец, следует отметить, что для дифференциального восстановления номер наиболее высокого используемого уровня  $N_B$  может только увеличиваться в независимых пакетах данных кадра,

тогда как уменьшение возможно в каждом кадре.

Подразумевается, что предложенный способ многоуровневого кодирования сжатого представления звука может быть реализован кодером для многоуровневого кодирования сжатого представления звука. Такой кодер может содержать соответствующие блоки, выполненные с возможностью выполнять соответствующие описанные выше этапы. Пример такого кодера 5000 схематично проиллюстрирован на фиг. 5. Например, такой кодер 5000 может содержать блок 5010 подразделения компонентов, выполненный с возможностью выполнять упомянутый выше этап S1010, блок 5020 присвоения компонентов, выполненный с возможностью выполнять упомянутый выше этап S1020, блок 5030 присвоения базовой вспомогательной информации, выполненный с возможностью выполнять упомянутый выше этап S1030, блок 5040 разбиения улучшающей вспомогательной информации, выполненный с возможностью выполнять упомянутый выше этап S1040, и блок 5050 присвоения улучшающей вспомогательной информации, выполненный с возможностью выполнять упомянутый выше этап S1050. Далее подразумевается, что соответствующие блоки такого кодера могут быть воплощены посредством процессора 5100 вычислительного устройства, который выполнено с возможностью выполнять обработку, выполняемую каждым из упомянутых соответствующих блоков, т.е., он выполнен с возможностью выполнять некоторые или все упомянутые выше этапы, а также любые дополнительные этапы предложенного метода кодирования. Кодер или вычислительное устройство могут дополнительно содержать память 5200, к которой процессор 5100 может осуществлять доступ.

Далее подразумевается, что предложенный способ декодирования сжатого представления звука, которое закодировано в множестве иерархических уровней, может быть реализован декодером для декодирования сжатого представления звука, которое закодировано в множестве иерархических уровней. Такой декодер может содержать соответствующие блоки, выполненные с возможностью выполнять соответствующие описанные выше этапы. Пример такого декодера 6000 схематично проиллюстрирован на фиг. 6. Например, такой декодер 6000 может содержать блок 6010

приема, выполненный с возможностью выполнять упомянутый выше этап S3010, блок 6020 определения первого индекса уровня, выполненный с возможностью выполнять упомянутый выше этап S3020, блок 6030 базового воссоздания, выполненный с возможностью выполнять упомянутый выше этап S3030, блок 6040 определения второго индекса уровня, выполненный с возможностью выполнять упомянутый выше этап S3040, и блок 6050, выполненный с возможностью выполнять упомянутый выше этап S3050. Далее подразумевается, что соответствующие блоки такого кодера могут быть воплощены посредством процессора 6100 вычислительного устройства, который выполнен с возможностью выполнять обработку, выполняемую каждым из упомянутых соответствующих блоков, т.е., он выполнен с возможностью выполнять некоторые или все упомянутые выше этапы, а также любые дополнительные этапы предложенного метода кодирования. Кодер или вычислительное устройство могут дополнительно содержать память 6200, к которой процессор 6100 может осуществлять доступ.

Следует отметить, что описание и чертежи лишь иллюстрируют принципы предложенных способов и устройств. Таким образом, будет очевидно, что специалисты в области техники смогут создавать различные структуры, которые, хотя явно не описаны и не показаны в настоящем документе, воплощают принципы изобретения и включены в пределы его сущности и объема. Кроме того, все примеры, приведенные в настоящем документе, преимущественно явно предназначены лишь для обучения, чтобы помочь читателю в понимании принципов предложенных способов и устройств, и концепций, внесенных изобретателями в развитие области техники, и должны быть истолкованы как не являющиеся ограничениями для таких специальным образом приведенных примеров и условий. Кроме того, предполагается, что все утверждения в настоящем документе, излагающие принципы, аспекты и варианты осуществления изобретения, а также их конкретные примеры, охватывают его эквиваленты.

Способы и устройство, описанные в настоящем документе, могут быть реализованы как программное обеспечение, программно-аппаратное обеспечение и/или аппаратные средства. Некоторые

компоненты, например, могут быть реализованы как программное обеспечение, работающее на процессоре цифровой обработки сигналов или микропроцессоре. Другие компоненты, например, могут быть реализованы как аппаратные средства и/или как специализированные интегральные схемы. Сигналы, встречающиеся в описанных способах и устройстве, могут быть сохранены на носителях, таких как оперативное запоминающее устройство или оптические запоминающие носители. Они могут быть перенесены через сети, такие как радиосети, спутниковые сети, беспроводные сети или проводные сети, например, Интернет.

Цитированная литература 1: ISO/IEC JTC1/SC29/WG11 23008-3:2015(E). Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, February 2015.

Цитированная литература 2: ISO/IEC JTC1/SC29/WG11 23008-3:2015/PDAM3. Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, AMENDMENT 3: MPEG-H 3D Audio Phase 2, July 2015.



**ФОРМУЛА ИЗОБРЕТЕНИЯ**

1. Способ декодирования сжатого представления звука или звукового поля системы Амбисоник высшего порядка (НОА), причем способ содержит этапы, на которых:

принимают битовый поток, содержащий сжатое представление НОА, соответствующее множеству иерархических уровней, которые включают в себя базовый уровень и один или более иерархических улучшающих уровней, причем множество уровней имеют присвоенные им компоненты базового сжатого представления звука или звукового поля, причем компоненты присваиваются соответствующим уровням в соответствующих группах компонентов,

декодируют сжатое представление НОА на основе базовой вспомогательной информации, которая связана с базовым уровнем, и на основе улучшающей вспомогательной информации, которая связана с одним или более иерархическими улучшающими уровнями,

причем базовая вспомогательная информация включает в себя базовую независимую вспомогательную информацию, относящуюся к первым индивидуальным монауральными сигналам, которые будут декодированы независимо от других монауральных сигналов.

2. Способ по п. 1, в котором базовая независимая вспомогательная информация указывает, что первые индивидуальные монауральные сигналы представляют направленный сигнал с направлением падения.

3. Способ по любому из пп. 1-2, в котором базовая вспомогательная информация дополнительно включает в себя базовую зависимую вспомогательную информацию, относящуюся к вторым индивидуальным монауральным сигналам, которые будут декодированы независимо от других монауральных сигналов.

4. Способ по п. 3, в котором базовая зависимая вспомогательная информация включает в себя основанные на векторах сигналы, которые распределены по направлениям в звуковом поле, причем распределение по направлениям определено посредством вектора.

5. Способ по п. 4, в котором компоненты вектора установлены равными нулю и не являются частью сжатого векторного представления.

6. Способ по любому из пп. 1-5, в котором компоненты базового сжатого представления звука соответствуют монауральным сигналам; и

монауральные сигналы представляют либо преобладающие звуковые сигналы, либо последовательности коэффициентов представления НОА.

7. Способ по любому из пп. 1-6, в котором битовый поток включает в себя полезные нагрузки данных, соответственно связанные с множеством иерархических уровней.

8. Способ по любому из пп. 1-7, в котором улучшающая вспомогательная информация включает в себя параметры, относящиеся по меньшей мере к одному из перечисленного: пространственное предсказание, синтез направленных подполосных сигналов и параметрическое дублирование звукового окружения.

9. Способ по любому из пп. 1-8, в котором улучшающая вспомогательная информация включает в себя информацию, которая обеспечивает возможность предсказания недостающих частей звука или звукового поля из направленных сигналов.

10. Способ по любому из пп. 1-9, дополнительно содержащий этапы, на которых:

определяют для каждого уровня, был ли соответствующий уровень принят корректно; и

определяют индекс уровня, находящегося непосредственно ниже наиболее низкого уровня, который не был принят корректно.

11. Устройство для декодирования сжатого представления звука или звукового поля системы Амбисоник высшего порядка (НОА), причем устройство содержит:

приемник для приема битового потока, содержащего сжатое представление НОА, соответствующее множеству иерархических уровней, которые включают в себя базовый уровень и один или более иерархических улучшающих уровней, причем множество уровней имеют присвоенные им компоненты базового сжатого представления звука или звукового поля, причем компоненты присваиваются соответствующим уровням в соответствующих группах компонентов,

декодер для декодирования сжатого представления НОА на основе базовой вспомогательной информации, которая связана с

базовым уровнем, и на основе улучшающей вспомогательной информации, которая связана с одним или более иерархическими улучшающими уровнями,

причем базовая вспомогательная информация включает в себя базовую независимую вспомогательную информацию, относящуюся к первым индивидуальным монауральными сигналам, которые будут декодированы независимо от других монауральных сигналов.

12. Устройство по п. 11, в котором базовая независимая вспомогательная информация включает в себя определение по меньшей мере монаурального сигнала для представления направленного сигнала с направлением падения.

13. Устройство по любому из пп. 11-12, в котором базовая вспомогательная информация дополнительно включает в себя базовую зависимую вспомогательную информацию, относящуюся к вторым индивидуальным монауральным сигналам, которые будут декодированы зависимо от других монауральных сигналов.

14. Устройство по п. 13, в котором базовая зависимая вспомогательная информация включает в себя основанные на векторах сигналы, которые распределены по направлениям в звуковом поле, причем распределение по направлениям определено посредством вектора.

15. Устройство по п. 14, в котором компоненты вектора установлены равными нулю и не являются частью сжатого векторного представления.

16. Устройство по любому из пп. 11-15, в котором компоненты базового сжатого представления звука соответствуют монауральным сигналам; и

монауральные сигналы представляют либо преобладающие звуковые сигналы, либо последовательности коэффициентов представления НОА.

17. Устройство по любому из пп. 11-16, в котором битовый поток включает в себя полезные нагрузки данных, соответственно связанные с множеством иерархических уровней.

18. Устройство по любому из пп. 11-17, в котором улучшающая вспомогательная информация включает в себя параметры, относящиеся по меньшей мере к одному из перечисленного:

пространственное предсказание, синтез направленных подполосных сигналов и параметрическое дублирование звукового окружения.

19. Устройство по любому из пп. 11-18, в котором улучшающая вспомогательная информация включает в себя информацию, которая обеспечивает возможность предсказания недостающих частей звука или звукового поля из направленных сигналов.

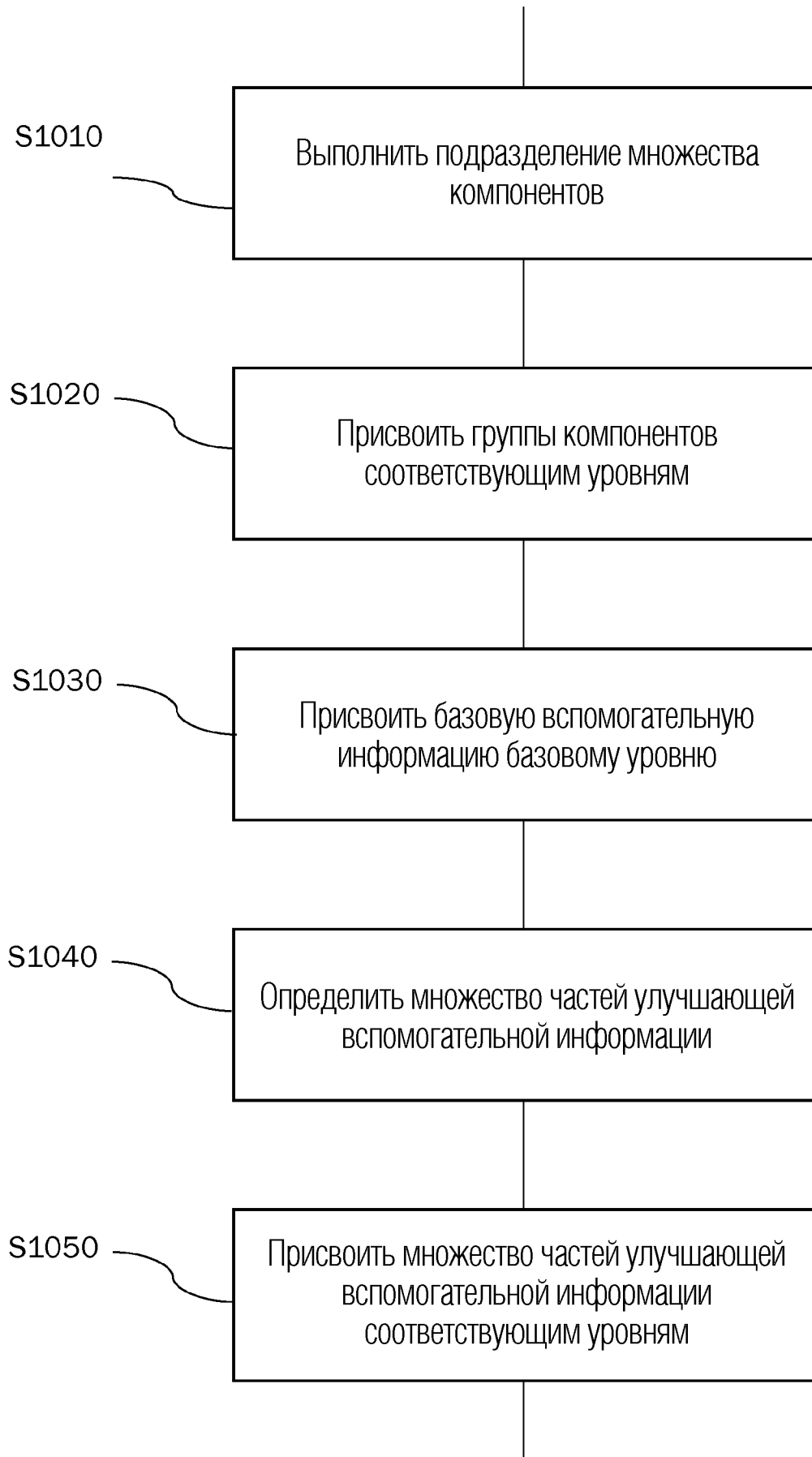
20. Устройство по любому из пп. 11-19, дополнительно содержащее:

определение для каждого уровня, был ли соответствующий уровень принят корректно; и

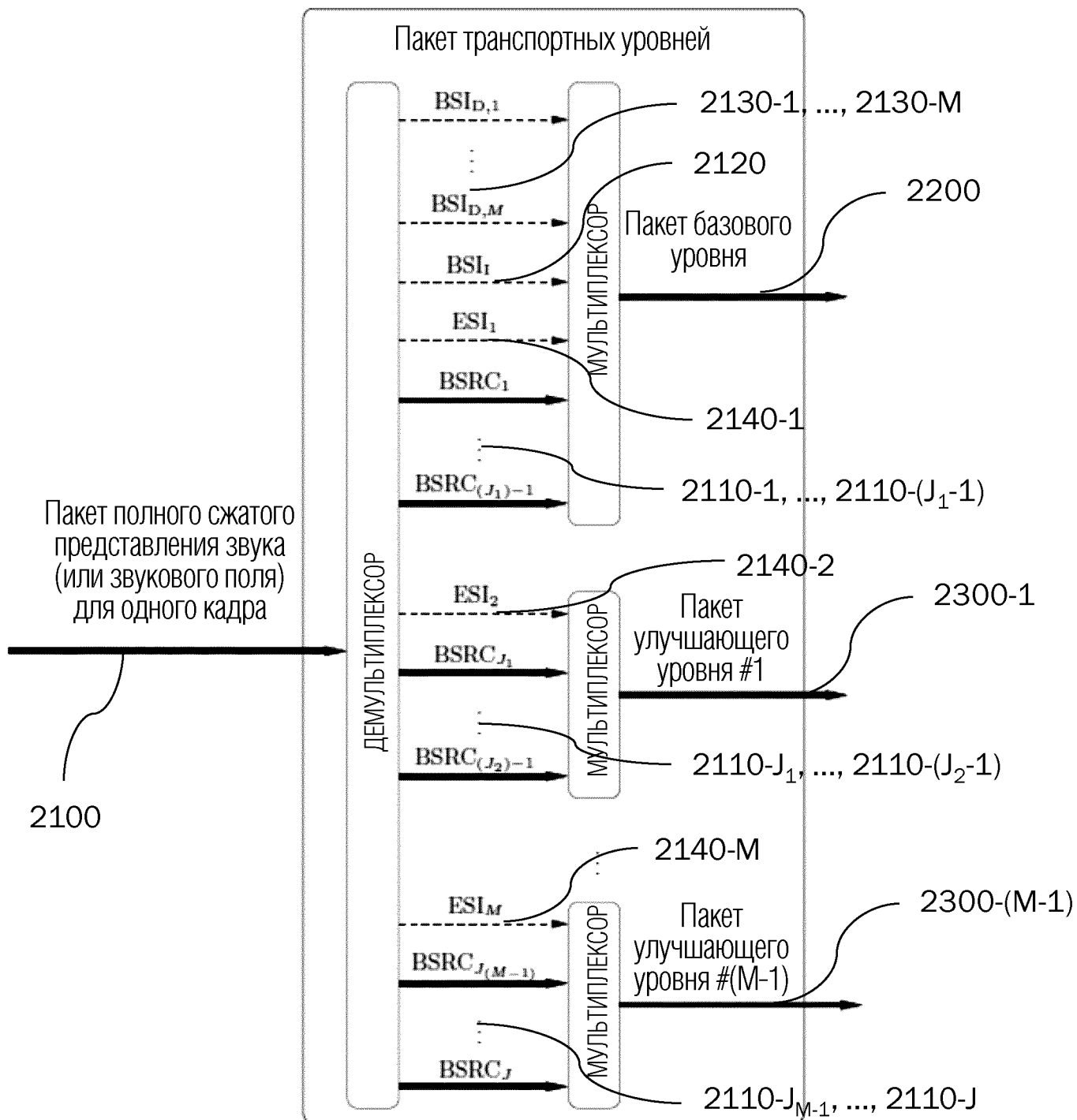
определение индекса уровня, находящегося непосредственно ниже наиболее низкого уровня, который не был принят корректно.

По доверенности

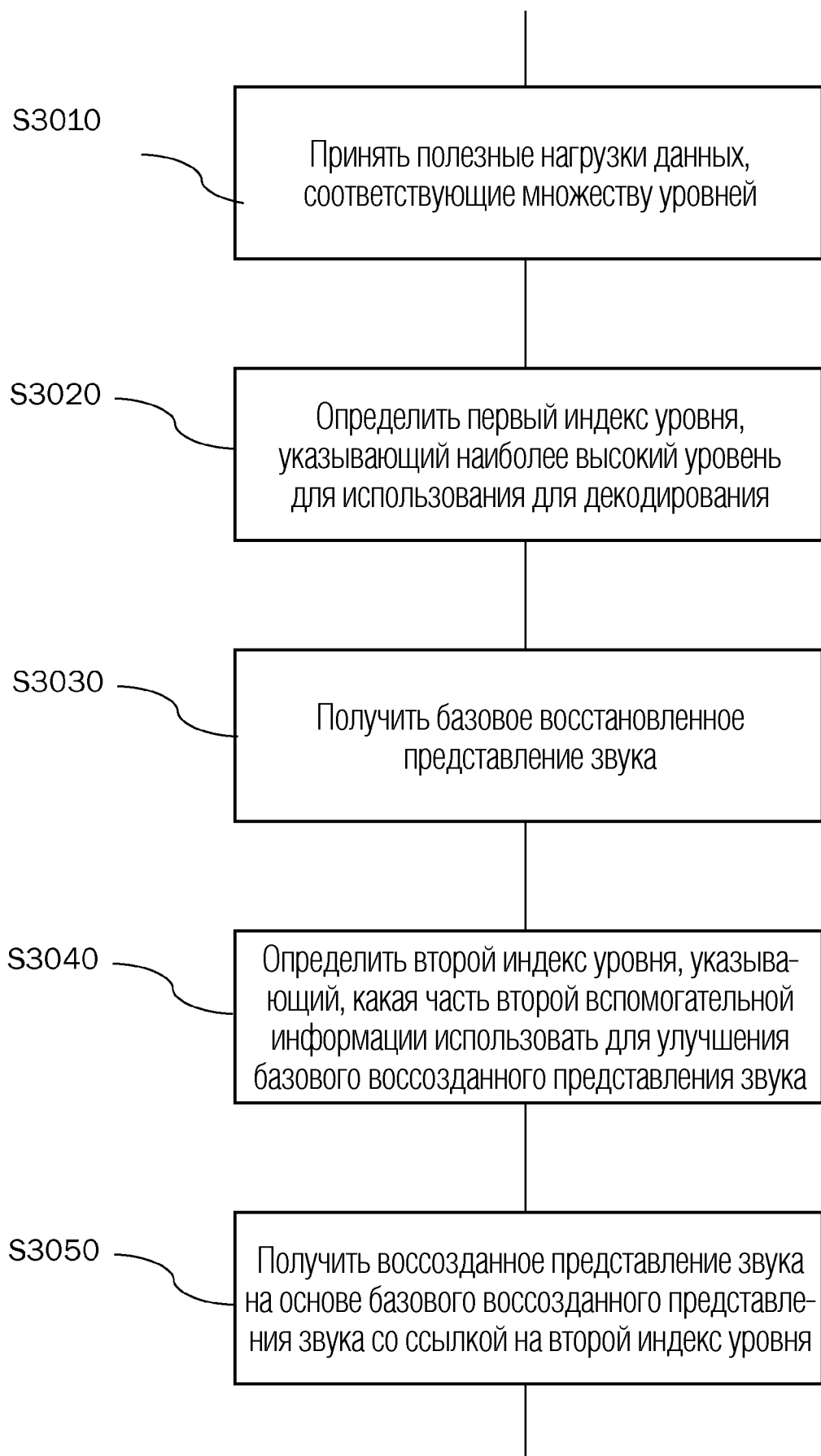
1/7



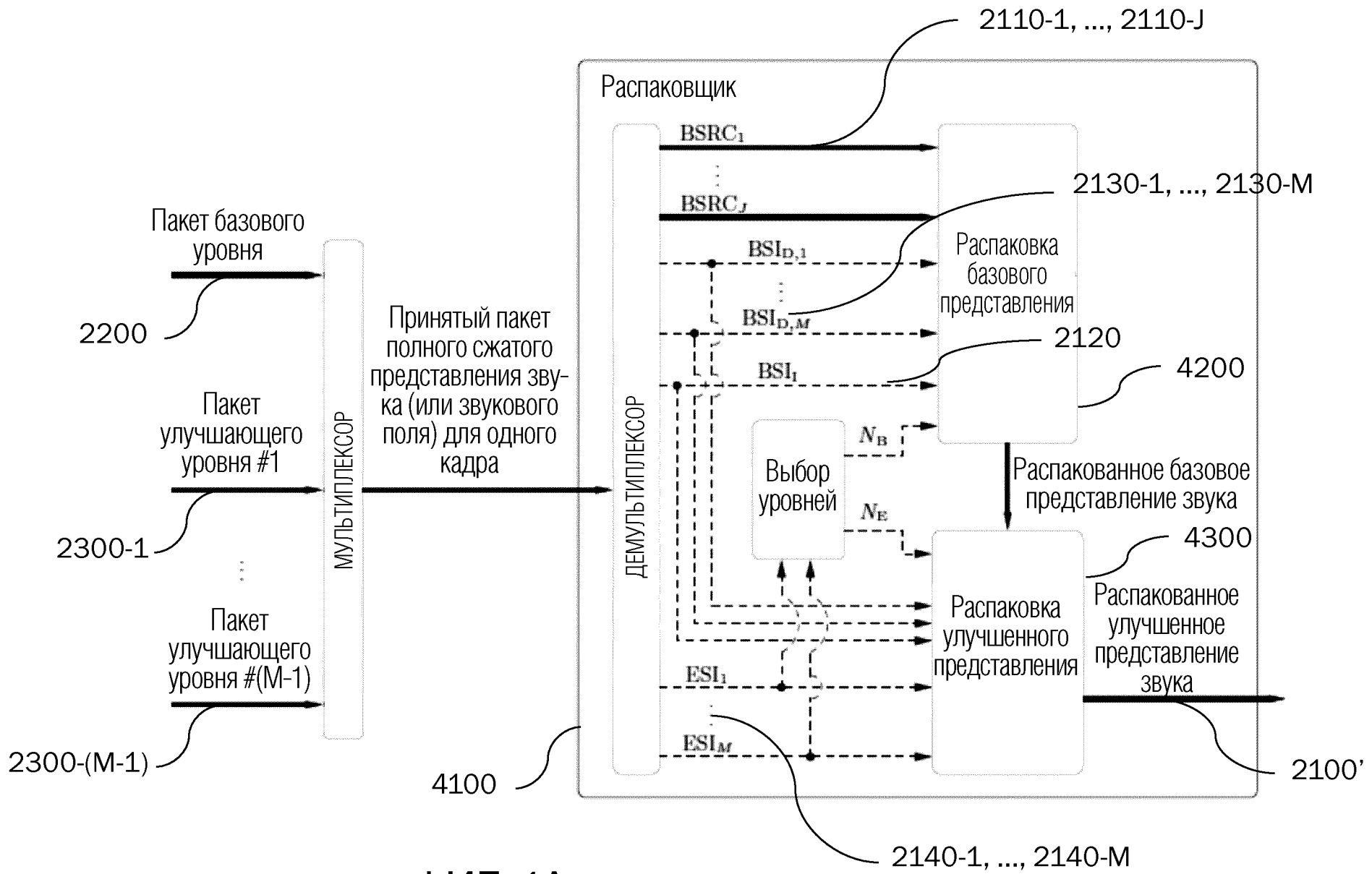
ФИГ. 1



ФИГ. 2

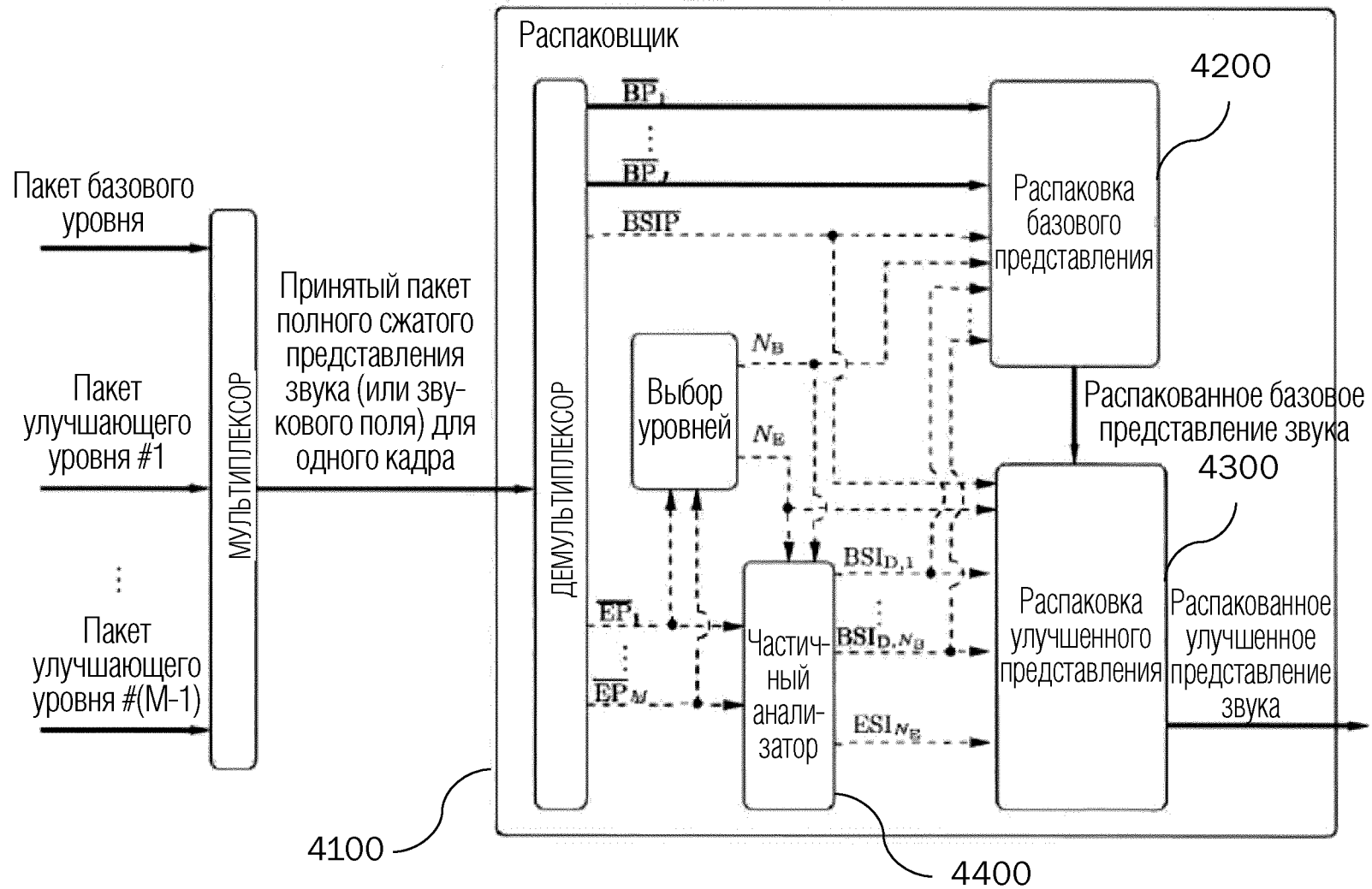


ФИГ. 3

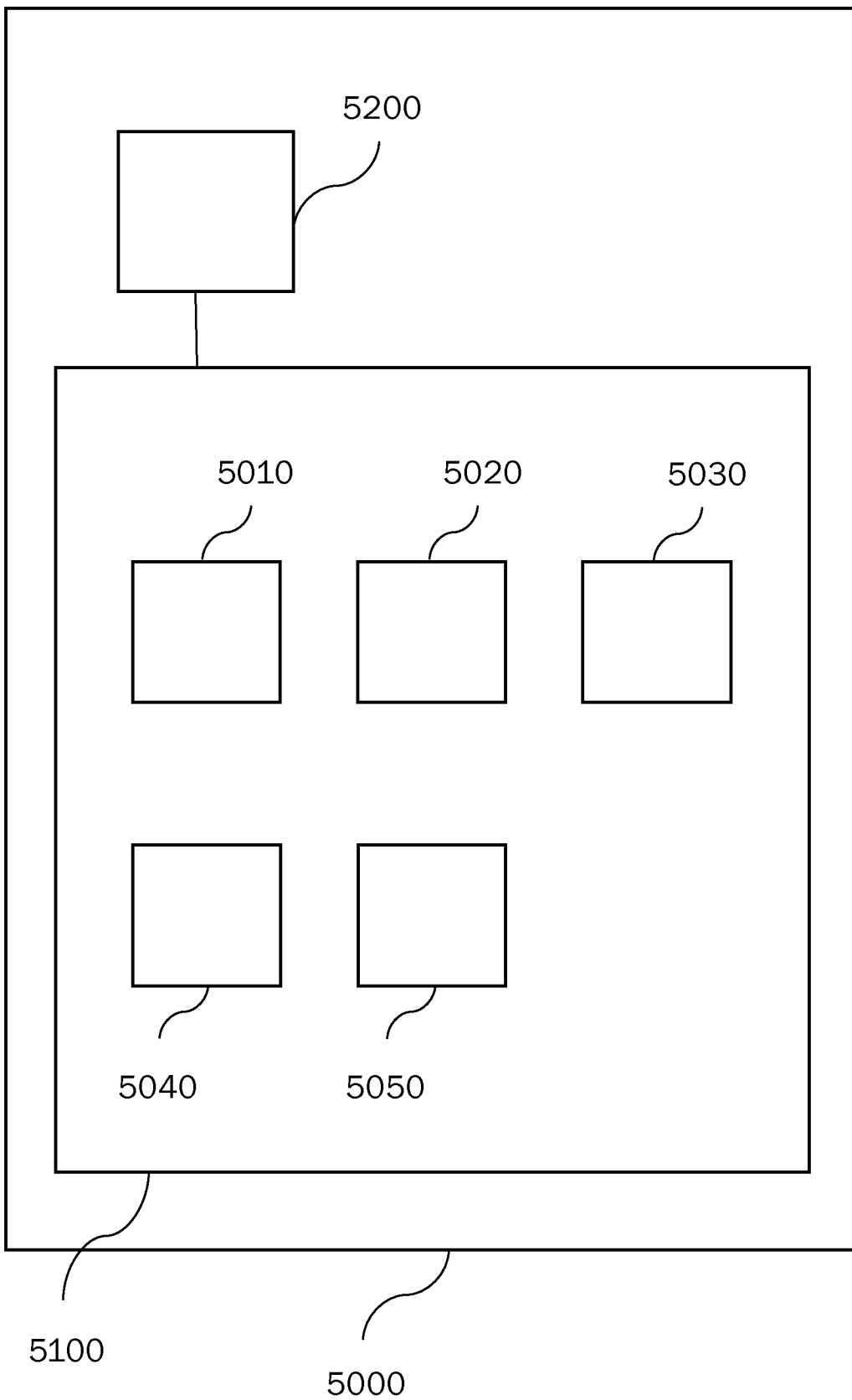


ФИГ. 4А

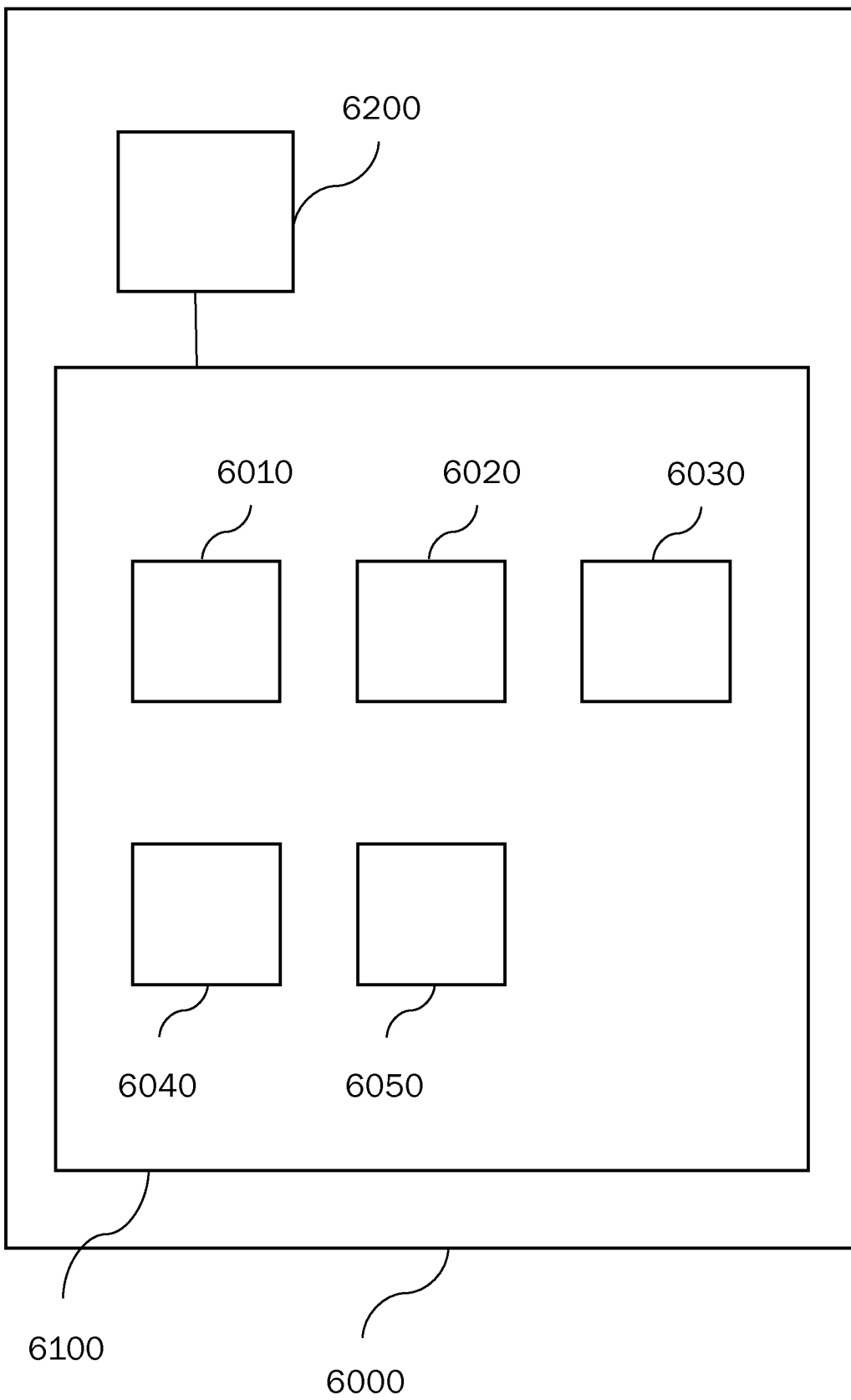




ФИГ. 4В



ФИГ. 5



ФИГ. 6