

**УКРАЇНА****(19) UA****(11) 123399****(13) C2****(51) МПК****G10L 19/008 (2013.01)**

**НАЦІОНАЛЬНИЙ ОРГАН
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ
ДЕРЖАВНЕ ПІДПРИЄМСТВО
"УКРАЇНСЬКИЙ ІНСТИТУТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ"**

(12) ОПИС ДО ПАТЕНТУ НА ВИНАХІД

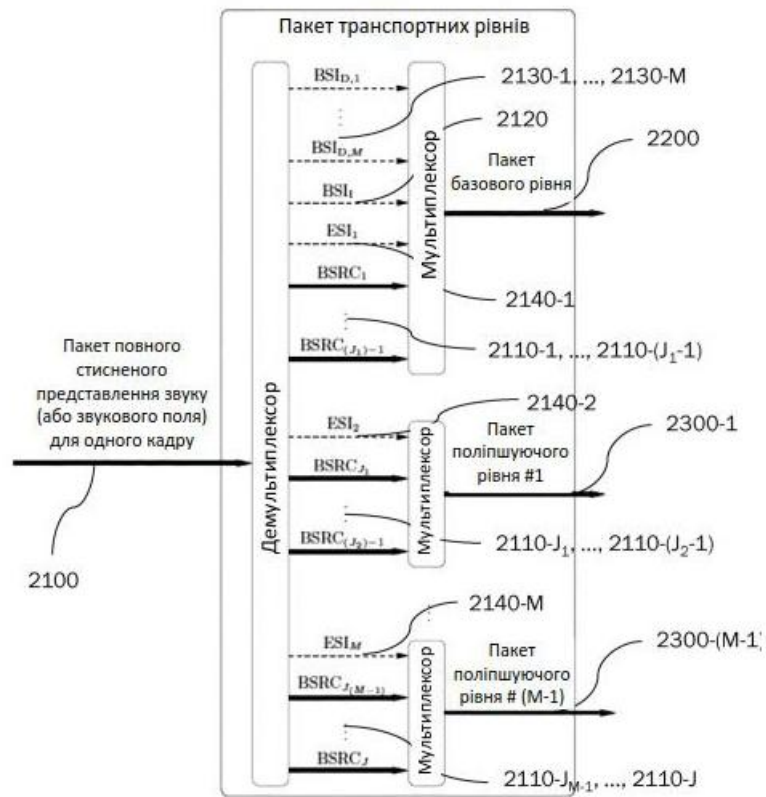
(21) Номер заявки:	а 2018 04907	(72) Винахідник(и):	Кордон Свен (DE), Крюгер Александр (DE)
(22) Дата подання заявки:	07.10.2016	(73) Володілець (володільці):	ДОЛБІ ІНТЕРНЕТНЛ АБ, Apollo Building, 3E Herikerbergweg 1-35, 1101 CN Amsterdam Zuidoost, The Netherlands (NL)
(24) Дата, з якої є чинними права інтелектуальної власності:	01.04.2021	(74) Представник:	Бочаров Максим Анатолійович
(31) Номер попередньої заявки відповідно до Паризької конвенції:	15306589.1, 15306653.5, 62/361,461, 62/361,416	(56) Перелік документів, взятих до уваги експертизою:	WO 2015140293 A1, 24.09.2015
(32) Дата подання попередньої заявки відповідно до Паризької конвенції:	08.10.2015, 15.10.2015, 12.07.2016, 12.07.2016		
(33) Код держави-учасниці Паризької конвенції, до якої подано попередню заявку:	EP, EP, US, US		
(41) Публікація відомостей про заявку:	25.09.2018, Бюл.№ 18		
(46) Публікація відомостей про державну реєстрацію:	31.03.2021, Бюл.№ 13		
(86) Номер та дата подання міжнародної заявки, поданої відповідно до Договору РСТ	PCT/EP2016/073969, 07.10.2016		

(54) БАГАТОРІВНЕВЕ КОДУВАННЯ СТИСНЕНИХ ПРЕДСТАВЛЕНЬ ЗВУКУ АБО ЗВУКОВОГО ПОЛЯ**(57) Реферат:**

Даний документ стосується способу багаторівневого кодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку включає базове стиснене представлення звуку, що містить множину компонентів, базову допоміжну інформацію для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля і поліпшуючу допоміжну інформацію, яка включає в себе параметри для поліпшення базового відтвореного представлення звуку. Спосіб включає підрозділяння множини компонентів на множину груп компонентів і присвоєння кожній з множини груп відповідному одному з множини ієрархічних рівнів, кількість груп відповідає кількості рівнів, і множина рівнів включає в себе базовий рівень і один або більше ієрархічних поліпшуючих рівнів, додавання базової допоміжної інформації до базового рівня, і визначення множини частин поліпшуючої допоміжної інформації з поліпшуючої допоміжної інформації і присвоєння кожній з множини частин поліпшуючої допоміжної інформації відповідному одному з множини рівнів, причому кожна частина поліпшуючої допоміжної інформації включає в себе параметри для поліпшення відтвореного представлення звуку, доступні з даних, включених у відповідний рівень і будь-які рівні нижче відповідного рівня. Документ також стосується способу

UA 123399 C2

декодування стисненого представлення звуку або звукового поля, причому стиснене представлення звуку закодоване в множині ієрархічних рівнів, які включають в себе базовий рівень і один або більше ієрархічних поліпшуючих рівнів, а також кодера і декодера для багаторівневого кодування стисненого представлення звуку.



Фіг. 2

ПЕРЕХРЕСНЕ ПОСИЛАННЯ НА СПОРІДНЕНІ ЗАЯВКИ

За даною заявкою запитується пріоритет відповідно до європейської патентної заявки № 15306590.9, поданої 8 жовтня 2015 року, і патентної заявки США № 62/361,809, зміст яких повністю включений в дану заявку за допомогою посилання.

5 ГАЛУЗЬ ТЕХНІКИ, ДО ЯКОЇ НАЛЕЖИТЬ ВІНАХІД

Даний документ стосується способів і пристроїв для багаторівневого аудіокодування. Зокрема, даний документ стосується способів і пристроїв для багаторівневого аудіокодування стиснених представлень звуку (або звукового поля), наприклад представлення звуку (або звукового поля) системи Амбісонік вищого порядку (Higher-Order Ambisonics, HOA).

10 РІВЕНЬ ТЕХНІКИ

Для потокового представлення звуку (або звукового поля) по каналу передачі зі змінними за часом умовами багаторівневе кодування є засобом, щоб адаптувати якість прийнятого представлення звуку до умов передачі і, зокрема, уникнути небажаних зникнень сигналу.

15 Для багаторівневого кодування представлення звуку (або звукового поля) звичайно підрозділяється на високопріоритетний базовий рівень відносно невеликого розміру і додаткові поліпшуючі рівні зі спадними пріоритетами і довільними розмірами. Кожен поліпшуючий рівень, як звичайно передбачається, містить наростаючу інформацію, щоб доповнити всі з більш низьких рівнів для поліпшення якості представлення звуку (або звукового поля). Величиною захисту від помилок для передачі окремих рівнів керують на основі їх пріоритету. Зокрема, базовому рівню надається високий захист від помилок, який є розумним і прийнятним внаслідок його малого розміру.

20 Однак існує потреба в багаторівневих схемах кодування для (розширеної версії) спеціальних типів стиснених представлень звуку або звукових полів, таких як, наприклад, стиснені представлення HOA звуку або звукового поля.

25 Даний документ вирішує згадані вище проблеми. Зокрема, описані способи і кодери/декодери для багаторівневого кодування стиснених представлень звуку або звукового поля.

СУТЬ ВІНАХОДУ

30 Відповідно до аспекту описаний спосіб багаторівневого кодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може включати в себе базове стиснене представлення звуку, що включає в себе множину компонентів. Множина компонентів може бути взаємодоповнюючими компонентами. Стиснене представлення звуку може додатково включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля.

35 Стиснене представлення звуку може, крім того, включати в себе поліпшуючу допоміжну інформацію, що включає в себе параметри для поліпшення (наприклад, розширення) базового відтвореного представлення звуку. Спосіб може включати в себе підрозділення (наприклад, угруповання) множини компонентів на множини груп компонентів. Спосіб може додатково включати в себе присвоєння (наприклад, додавання) кожної з множини груп відповідному

40 одному з множини ієрархічних рівнів. Присвоєння може вказувати відповідність між відповідними групами і рівнями. Можна сказати, що компоненти, присвоєні відповідному рівню, включені в цей рівень. Кількість груп може відповідати (наприклад, дорівнювати) кількості рівнів. Множина рівнів може включати в себе базовий рівень і один або більше ієрархічних поліпшуючих рівнів. Множина ієрархічних рівнів може бути впорядкована від базового рівня, через перший поліпшуючий рівень, другий поліпшуючий рівень і т. д., аж до загального найбільш високого поліпшуючого рівня (загального найбільш високого рівня). Спосіб може

45 додатково включати в себе додавання базової допоміжної інформації до базового рівня (наприклад, включення базової допоміжної інформації в базовий рівень або розподіл базової допоміжної інформації базовому рівню, наприклад, з метою передачі або зберігання). Спосіб

50 може додатково включати в себе визначення множини частин поліпшуючої допоміжної інформації на основі поліпшуючої допоміжної інформації. Спосіб може, крім того, включати в себе присвоєння (наприклад, додавання) кожної з множини частин поліпшуючої допоміжної інформації відповідному одному з множини рівнів. Кожна частина поліпшуючої допоміжної інформації може включати в себе параметри для поліпшення відтвореного (наприклад, відновленого) представлення звуку, доступні з даних, включених (наприклад, присвоєних або доданих) у відповідний рівень і будь-які рівні нижче відповідного рівня. Багаторівневе кодування може бути виконане з метою передачі по каналу передачі або з метою збереження на придатному запам'ятовувальному носії, такому як, наприклад, CD, DVD або Blu-ray Disc™.

60 Сконфігурований, як згадано вище, запропонований спосіб дозволяє ефективно застосовувати багаторівневе кодування до стиснених представлень звуку, що містять множину

компонентів, а також першу і поліпшуючу допоміжну інформацію (наприклад, незалежну базову допоміжну інформацію і поліпшуючу допоміжну інформацію), що мають викладені вище властивості. Зокрема, запропонований спосіб гарантує, що кожен рівень включає в себе придатну допоміжну інформацію для відновлення відтвореного представлення звуку з компонентів, включених у будь-які рівні аж до рівня, що розглядається. При цьому передбачається, що рівні аж до розглянутого рівня включають в себе, наприклад, базовий рівень, перший поліпшуючий рівень, другий поліпшуючий рівень і т. д., аж до рівня, що розглядається. Таким чином, незалежно від фактичного найбільш високого застосовного рівня (наприклад, рівня нижче найбільш низького шару, який не був коректно прийнятий, і, таким чином, всі рівні нижче найбільш високого застосовного рівня і сам найбільш високий застосовний рівень прийняті коректно), декодеру дозволяється поліпшити або розширити відтворене представлення звуку, навіть якщо відтворене представлення звуку може відрізнитися від повного представлення звуку. Зокрема, незалежно від фактичного найбільш високого застосовного рівня, для декодера достатньо декодувати корисне навантаження поліпшуючої допоміжної інформації тільки для єдиного шару (тобто для найбільш високого застосовного рівня), щоб поліпшити або розширити відтворене представлення звуку, яке доступне на основі всіх компонентів, включених у рівні аж до фактичного найбільш високого застосовного рівня. Таким чином, для кожного часового інтервалу (наприклад, кадру) повинно бути декодоване тільки єдине корисне навантаження поліпшуючої допоміжної інформації. З іншого боку, запропонований спосіб дозволяє повністю використовувати перевагу скорочення необхідної ширини смуги, що може бути досягнуто при застосуванні багаторівневого кодування.

У варіантах здійснення компоненти базового стисненого представлення звуку можуть відповідати монауральним сигналам (наприклад, транспортним сигналам або монауральним транспортним сигналам). Монауральні сигнали можуть представляти або переважні звукові сигнали, або послідовності коефіцієнтів представлення НОА. Монауральні сигнали можуть бути квантовані.

У варіантах здійснення базова допоміжна інформація може включати в себе інформацію, яка визначає декодування (наприклад, відновлення) одного або більше з множини компонентів індивідуально, незалежно від інших компонентів. Наприклад, базова допоміжна інформація може представляти допоміжну інформацію, яка стосується індивідуальних монауральних сигналів, незалежно від інших монауральних сигналів. Таким чином, базова допоміжна інформація може згадуватися як незалежна базова допоміжна інформація.

У варіантах здійснення поліпшуюча допоміжна інформація може представляти поліпшуючу допоміжну інформацію. Поліпшуюча допоміжна інформація може включати в себе параметри прогнозування для базового стисненого представлення звуку для поліпшення (наприклад, розширення) базового відтвореного представлення звуку, які доступні з базового стисненого представлення звуку і базової допоміжної інформації.

У варіантах здійснення спосіб може додатково включати в себе формування транспортного потоку для передачі даних множини рівнів (наприклад, даних, присвоєних або доданих до відповідних рівнів або іншим чином включених у відповідні рівні). Базовий рівень може мати найбільш високий пріоритет передачі, і ієрархічні поліпшуючі рівні можуть мати спадні пріоритети передачі. Таким чином, пріоритет передачі може зменшуватися від базового рівня до першого поліпшуючого рівня, від першого поліпшуючого рівня до другого поліпшуючого рівня і т. д. Величиною захисту від помилок для передачі даних множини рівнів можна керувати відповідно до відповідних пріоритетів передачі. Тим самим може бути забезпечено, що щонайменше деяка кількість більш низьких рівнів передається достовірно, з іншого боку, скорочуючи повну необхідну ширину смуги без застосування надмірного захисту від помилок до більш високих рівнів.

У варіантах здійснення спосіб може додатково включати в себе для кожного з множини рівнів формування пакета транспортного рівня, що включає в себе дані відповідного рівня. Наприклад, для кожного часового інтервалу (наприклад, кадру) відповідний пакет транспортного рівня може бути сформований для кожного з множини рівнів.

У варіантах здійснення стиснене представлення звуку може додатково включати в себе додаткову базову допоміжну інформацію для декодування базового стисненого представлення звуку в базове відтворене представлення звуку. Додаткова базова допоміжна інформація може включати в себе інформацію, яка визначає декодування одного або більше з множини компонентів залежно від відповідних інших компонентів. Спосіб може додатково включати в себе виконання декомпозиції додаткової базової допоміжної інформації на множини частин додаткової базової допоміжної інформації. Спосіб може, крім того, включати в себе додавання частин додаткової базової допоміжної інформації до базового рівня (наприклад, включення

частини додаткової базової допоміжної інформації в базовий рівень або розподіл частин додаткової базової допоміжної інформації базовому рівню, наприклад, з метою передачі або зберігання). Кожна частина додаткової базової допоміжної інформації може бути пов'язана з відповідним рівнем і може включати в себе інформацію, яка визначає декодування одного або

5 більше компонентів, присвоєних відповідному рівню, залежно (тільки) від відповідних інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня. Таким чином, кожна частина додаткової базової допоміжної інформації визначає компоненти на відповідному рівні, якому відповідає ця частина додаткової базової допоміжної інформації, незалежно від будь-яких інших компонентів, присвоєних більш високим рівням, ніж відповідний рівень.

Сконфігурований таким чином запропонований спосіб уникає фрагментації додаткової базової допоміжної інформації за допомогою додавання всіх частин до базового рівня. Інакше кажучи, всі частини додаткової базової допоміжної інформації включені в базовий рівень. Декомпозиція додаткової базової допоміжної інформації гарантує, що для кожного рівня

15 доступна частина додаткової базової допоміжної інформації, що не вимагає знання компонентів на більш високих рівнях. Таким чином, незалежно від фактичного найбільш високого застосовного рівня, для декодера достатньо декодувати додаткову базову допоміжну інформацію, включену в рівні аж до найбільш високого застосовного рівня.

У варіантах здійснення додаткова базова допоміжна інформація може включати в себе інформацію, яка визначає декодування (наприклад, відновлення) одного або більше з множини компонентів залежно від інших компонентів. Наприклад, додаткова базова допоміжна інформація може представляти допоміжну інформацію, яка стосується індивідуальних монауральних сигналів, залежно від інших монауральних сигналів. Таким чином, додаткова базова допоміжна інформація може згадуватися як залежна базова допоміжна інформація.

У варіантах здійснення стиснене представлення звуку може бути оброблене для послідовних часових інтервалів, наприклад часових інтервалів однакового розміру. Послідовні часові інтервали можуть бути кадрами. Таким чином, спосіб може працювати на основі кадрів, тобто стиснене представлення звуку може бути закодоване покадрово. Стиснене представлення звуку може бути доступне для кожного послідовного часового інтервалу (наприклад, для кожного кадру). Таким чином, операція стиснення, за допомогою якої було

20 одержане стиснене представлення звуку, може працювати на основі кадрів.

У варіантах здійснення спосіб може додатково включати в себе формування інформації конфігурації, яка вказує для кожного рівня компоненти базового стисненого представлення звуку, які присвоєні цьому рівню. Таким чином, декодер може без складностей здійснити доступ

25 до інформації, необхідної для декодування, без непотрібного аналізу прийнятих корисних навантажень даних.

Відповідно до іншого аспекту описаний спосіб багаторівневого кодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може включати в себе базове стиснене представлення звуку, що включає в себе множину компонентів. Множина

30 компонентів може бути взаємодоповнюючими компонентами. Стиснене представлення звуку може додатково включати в себе базову допоміжну інформацію (наприклад, незалежну базову допоміжну інформацію) і третю інформацію (наприклад, залежну базову допоміжну інформацію) для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля. Базова допоміжна інформація може включати в себе інформацію, яка визначає декодування одного або більше з множини компонентів індивідуально, незалежно від інших компонентів. Додаткова базова допоміжна інформація може включати в себе інформацію, яка визначає декодування одного або більше з множини компонентів залежно від відповідних інших компонентів. Спосіб може включати в себе підрозділяння (наприклад, угруповання) множини компонентів на множину груп компонентів. Спосіб може додатково включати в себе

35 присвоєння (наприклад, додавання) кожної з множини груп відповідному одному з множини ієрархічних рівнів. Присвоєння може вказувати відповідність між відповідними групами і рівнями. Можна сказати, що компоненти, присвоєні відповідному рівню, включені в цей рівень. Кількість груп може відповідати (наприклад, дорівнювати) кількості рівнів. Множина рівнів може включати в себе базовий рівень і один або більше ієрархічних поліпшувачих рівнів. Спосіб може додатково

40 включати в себе додавання базової допоміжної інформації до базового рівня (наприклад, включення базової допоміжної інформації в базовий рівень або розподіл базової допоміжної інформації базовому рівню, наприклад, з метою передачі або зберігання). Спосіб може додатково включати в себе виконання декомпозиції додаткової базової допоміжної інформації на множину частин додаткової базової допоміжної інформації і додавання частин додаткової

45 базової допоміжної інформації до базового рівня (наприклад, включення частин додаткової

50

55

60

базової допоміжної інформації в базовий рівень або розподіл частин додаткової базової допоміжної інформації базовому рівню, наприклад, з метою передачі або зберігання). Кожна частина додаткової базової допоміжної інформації може бути пов'язана з відповідним рівнем і включати в себе інформацію, яка визначає декодування одного або більше компонентів, присвоєних відповідному рівню, залежно від відповідних інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня.

Сконфігурований таким чином запропонований спосіб гарантує, що для кожного рівня доступна придатна додаткова базова допоміжна інформація для декодування компонентів, включених у будь-який рівень аж до відповідного рівня, не вимагаючи коректного прийому або декодування (або, у цілому, знання) будь-яких більш високих рівнів. У випадку стисненого представлення НОА запропонований спосіб гарантує, що в режимі векторного кодування придатний V-вектор доступний для всього компонента, що належить рівням аж до найбільш високого застосовного рівня. Зокрема, запропонований спосіб виключає випадок, у якому елементи V-вектора, що відповідає компонентам на більш високих рівнях, явно не повідомлені. Відповідно до цього інформація, включена в рівні аж до найбільш високого застосовного рівня, є достатньою для декодування (наприклад, відновлення) будь-яких компонентів, що належать рівням аж до найбільш високого застосовного рівня. Тим самим забезпечується придатне відновлення відповідних відтворених представлень НОА для більш низьких рівнів, навіть якщо більш високі рівні не могли бути коректно прийняті декодером. З іншого боку, запропонований спосіб дозволяє повністю використовувати переваги скорочення необхідної ширини смуги, що може бути досягнуто при застосуванні багаторівневого кодування.

Варіанти здійснення цього аспекту можуть належати до варіантів здійснення попереднього аспекту.

Відповідно до іншого аспекту описаний спосіб декодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може бути закодоване у множині ієрархічних рівнів. Множина ієрархічних рівнів може включати в себе базовий рівень і один або декілька ієрархічних зростаючих рівнів. Множина рівнів може мати присвоєні їм компоненти базового стисненого представлення звуку або звукового поля. Інакше кажучи, множина рівнів може включати в себе компоненти базової стисненої допоміжної інформації. Компоненти можуть бути присвоєні відповідним рівням у відповідних групах компонентів. Множина компонентів може бути взаємодоповнюючими компонентами. Базовий рівень може включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку. Кожен рівень може включати в себе частину поліпшуючої допоміжної інформації, що включає в себе параметри для поліпшення базового відтвореного представлення звуку, доступні з даних, включених у відповідний рівень і будь-які рівні нижче відповідного рівня. Спосіб може включати в себе прийом корисних навантажень даних, відповідно пов'язаних з множиною ієрархічних рівнів. Спосіб може додатково включати в себе визначення першого індексу рівня, що вказує найбільш високий застосовний рівень з множини рівнів для використання для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля. Спосіб може додатково включати в себе одержання базового відтвореного представлення звуку з компонентів, присвоєних найбільш високому застосовному рівню і будь-яким рівням нижче найбільш високого застосовного рівня, з використанням базової допоміжної інформації. Спосіб може додатково включати в себе визначення другого індексу рівня, що вказує, яка частина поліпшуючої допоміжної інформації повинна використовуватися для поліпшення (наприклад, розширення) базового відтвореного представлення звуку. Спосіб, крім того, може включати в себе одержання відтвореного представлення звуку або звукового поля з базового відтвореного представлення звуку з посиланням на другий індекс рівня.

Сконфігурований таким чином запропонований спосіб гарантує, що відтворене представлення звуку має оптимальну якість з використанням доступної (наприклад, коректно прийнятої) інформації в найкращому можливому ступені.

У варіантах здійснення компоненти базового стисненого представлення звуку можуть відповідати монауральним сигналам (наприклад, монауральним транспортним сигналам). Монауральні сигнали можуть представляти або переважні звукові сигнали, або послідовності коефіцієнтів представлення НОА. Монауральні сигнали можуть бути квантовані.

У варіантах здійснення базова допоміжна інформація може включати в себе інформацію, яка визначає декодування (наприклад, відновлення) одного або більше з множини компонентів індивідуально, незалежно від інших компонентів. Наприклад, базова допоміжна інформація може представляти допоміжну інформацію, яка стосується індивідуальних монауральних сигналів, незалежно від інших монауральних сигналів. Таким чином, базова допоміжна інформація може згадуватися як незалежна базова допоміжна інформація.

У варіантах здійснення поліпшуюча допоміжна інформація може представляти поліпшуючу допоміжну інформацію. Поліпшуюча допоміжна інформація може включати в себе параметри прогнозування для базового стисненого представлення звуку для поліпшення (наприклад, розширення) базового відтвореного представлення звуку, які доступні з базового стисненого представлення звуку і базової допоміжної інформації.

У варіантах здійснення спосіб може додатково включати в себе визначення для кожного рівня, чи був відповідний рівень прийнятий коректно. Спосіб може додатково включати в себе визначення першого індексу рівня як індексу того рівня, який знаходиться безпосередньо нижче найбільш низького рівня, що не був прийнятий коректно.

У варіантах здійснення визначення другого індексу рівня може включати в себе або визначення другого індексу рівня як такого, що дорівнює першому індексу рівня, або визначення значення індексу як другого індексу рівня, яке вказує, що не слід використовувати яку-небудь поліпшуючу допоміжну інформацію при одержанні відтвореного представлення звуку. В останньому випадку відтворене представлення звуку може бути еквівалентне базовому відтвореному представленню звуку.

У варіантах здійснення корисні навантаження даних можуть бути прийняті і оброблені для послідовних часових інтервалів, наприклад часових інтервалів однакового розміру. Послідовні часові інтервали можуть бути кадрами. Таким чином, спосіб може працювати на основі кадрів. Спосіб може додатково включати в себе визначення другого індексу рівня як такого, що дорівнює першому індексу рівня, якщо стиснені представлення звуку для послідовних часових інтервалів можуть бути декодовані незалежно одне від одного.

У варіантах здійснення корисні навантаження даних можуть бути прийняті і оброблені для послідовних часових інтервалів, наприклад часових інтервалів однакового розміру. Послідовні часові інтервали можуть бути кадрами. Таким чином, спосіб може працювати на основі кадрів. Спосіб може додатково включати в себе для даного часового інтервалу з послідовних часових інтервалів визначення для кожного рівня, чи був відповідний рівень прийнятий коректно, якщо стиснені представлення звуку для послідовних часових інтервалів не можуть бути декодовані незалежно одне від одного. Спосіб може додатково включати в себе визначення першого індексу рівня для даного часового інтервалу як меншого індексу з першого індексу рівня часового інтервалу, що передує даному часовому інтервалу, і індексу рівня, який знаходиться безпосередньо нижче найбільш низького рівня, що не був прийнятий коректно.

У варіантах здійснення спосіб може додатково включати в себе для даного часового інтервалу визначення, чи дорівнює перший індекс рівня для даного часового інтервалу першому індексу рівня для попереднього часового інтервалу, якщо стиснені представлення звуку для послідовних часових інтервалів не можуть бути декодовані незалежно одне від одного. Спосіб може додатково включати в себе визначення, що другий індекс рівня для даного часового інтервалу дорівнює першому індексу рівня для даного часового інтервалу, якщо перший індекс рівня для даного часового інтервалу дорівнює першому індексу рівня для попереднього часового інтервалу. Спосіб може додатково включати в себе визначення значення індексу як другого індексу рівня, який вказує, що не слід використовувати яку-небудь поліпшуючу допоміжну інформацію при одержанні відтвореного представлення звуку, якщо перший індекс рівня для даного часового інтервалу не дорівнює першому індексу рівня для попереднього часового інтервалу.

У варіантах здійснення базовий рівень може включати в себе щонайменше одну частину додаткової базової допоміжної інформації, яка пов'язана з відповідним рівнем і включає в себе інформацію, що визначає декодування одного або більше компонентів з компонентів, присвоєних відповідному рівню, залежно від інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня. Спосіб може додатково включати в себе для кожної частини додаткової базової допоміжної інформації декодування частини додаткової базової допоміжної інформації за допомогою посилання на компоненти, присвоєні їй відповідному рівню і будь-яким рівням нижче відповідного рівня. Спосіб може додатково включати в себе корекцію частини додаткової базової допоміжної інформації за допомогою посилання на компоненти, присвоєні найбільш високому застосовному рівню і будь-яким рівням між найбільш високим застосовним рівнем і відповідним рівнем. Базове відтворене представлення звуку може бути одержане з компонентів, присвоєних найбільш високому застосовному рівню і будь-яким рівням нижче найбільш високого застосовного рівня, з використанням базової допоміжної інформації і скоректованої частини додаткової базової допоміжної інформації, одержаної із частин додаткової базової допоміжної інформації, що відповідають рівням аж до найбільш високого застосовного рівня.

У варіантах здійснення додаткова базова допоміжна інформація може включати в себе інформацію, яка визначає декодування (наприклад, відновлення) одного або більше з множини компонентів залежно від інших компонентів. Наприклад, додаткова базова допоміжна інформація може представляти допоміжну інформацію, яка стосується індивідуальних монауральних сигналів, залежно від інших монауральних сигналів. Таким чином, додаткова базова допоміжна інформація може згадуватися як залежна базова допоміжна інформація.

Відповідно до іншого аспекту описаний спосіб декодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може бути закодоване в множині ієрархічних рівнів. Множина ієрархічних рівнів може включати в себе базовий рівень і один або декілька ієрархічних зростаючих рівнів. Множина рівнів може мати присвоєні їм компоненти базового стисненого представлення звуку або звукового поля. Інакше кажучи, множина рівнів може включати в себе компоненти базової стисненої допоміжної інформації. Компоненти можуть бути присвоєні відповідним рівням у відповідних групах компонентів. Множина компонентів може бути взаємодоповнюючими компонентами. Базовий рівень може включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку. Базовий рівень може додатково включати в себе щонайменше одну частину додаткової базової допоміжної інформації, яка пов'язана з відповідним рівнем і включає в себе інформацію, що визначає декодування одного або більше компонентів з компонентів, присвоєних відповідному рівню, залежно від інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня. Спосіб може включати в себе прийом корисних навантажень даних, відповідно пов'язаних з множиною ієрархічних рівнів. Спосіб може додатково включати в себе визначення першого індексу рівня, що вказує найбільш високий застосовний рівень з множини рівнів для використання для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля. Спосіб може додатково включати в себе для кожної частини додаткової базової допоміжної інформації декодування частини додаткової базової допоміжної інформації за допомогою посилання на компоненти, присвоєні її відповідному рівню і будь-яким рівням нижче відповідного рівня. Спосіб може додатково включати в себе для кожної частини додаткової базової допоміжної інформації корекцію частини додаткової базової допоміжної інформації за допомогою посилання на компоненти, присвоєні найбільш високому застосовному рівню і будь-яким рівням між найбільш високим застосовним рівнем і відповідним рівнем. Базове відтворене представлення звуку може бути одержане з компонентів, присвоєних найбільш високому застосовному рівню і будь-яким рівням нижче найбільш високого застосовного рівня, з використанням базової допоміжної інформації і скоректованої частини додаткової базової допоміжної інформації, одержаної із частин додаткової базової допоміжної інформації, що відповідають рівням аж до найбільш високого застосовного рівня. Спосіб може додатково включати визначення другого індексу рівня, який або дорівнює першому індексу рівня, або вказує опущення поліпшуючої допоміжної інформації під час декодування.

Сконфігурований таким чином запропонований спосіб гарантує, що додаткова базова допоміжна інформація, яка в остаточному підсумку використовується для декодування базового стисненого представлення звуку, не включає в себе надлишкові елементи, тим самим реалізуючи більш ефективне фактичне декодування базового стисненого представлення звуку.

Варіанти здійснення цього аспекту можуть належати до варіантів здійснення попереднього аспекту.

Відповідно до іншого аспекту описаний кодер для багаторівневого кодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може включати в себе базове стиснене представлення звуку, що включає в себе множину компонентів. Множина компонентів може бути взаємодоповнюючими компонентами. Стиснене представлення звуку може додатково включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля. Стиснене представлення звуку може, крім того, включати в себе поліпшуючу допоміжну інформацію, яка включає в себе параметри для поліпшення (наприклад, розширення) базового відтвореного представлення звуку. Кодер може включати в себе процесор, виконаний з можливістю виконувати деякі або всі етапи способів відповідно до першого згаданого вище аспекту і другого згаданого вище аспекту.

Відповідно до іншого аспекту описаний декодер для декодування стисненого представлення звуку або звукового поля. Стиснене представлення звуку може бути закодоване в множині ієрархічних рівнів. Множина ієрархічних рівнів може включати в себе базовий рівень і один або декілька ієрархічних зростаючих рівнів. Множина рівнів може мати присвоєні їм компоненти базового стисненого представлення звуку або звукового поля. Інакше кажучи, множина рівнів

може включати в себе компоненти базової стисненої допоміжної інформації. Компоненти можуть бути присвоєні відповідним рівням у відповідних групах компонентів. Множина компонентів може бути взаємодоповнюючими компонентами. Базовий рівень може включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку.

5 Кожен рівень може включати в себе частину поліпшуючої допоміжної інформації, яка включає в себе параметри для поліпшення (наприклад, розширення) базового відтвореного представлення звуку, доступні з даних, включених у відповідний рівень і будь-які рівні нижче відповідного рівня. Декодер може включати в себе процесор, виконаний з можливістю виконувати деякі або всі етапи способів відповідно до третього згаданого вище аспекту і

10 четвертого згаданого вище аспекту.

Відповідно до інших аспектів способи, пристрої і системи спрямовані на декодування стисненого представлення Higher-Order Ambisonics (HOA) звуку або звукового поля (просторового звуку або звукового поля вищого порядку). Пристрій може мати приймач, виконаний з можливістю приймати, або спосіб може приймати бітовий потік, який містить

15 стиснене представлення HOA, що відповідає множині ієрархічних рівнів, які включають в себе базовий рівень і один або більше ієрархічних поліпшуючих рівнів. Множина рівнів має присвоєні їм компоненти базового стисненого представлення звуку або звукового поля, компоненти присвоєні відповідним рівням у відповідних групах компонентів. Пристрій може мати декодер, виконаний з можливістю декодувати, або спосіб може декодувати стиснене представлення HOA

20 на основі базової допоміжної інформації, яка пов'язана з базовим рівнем, і на основі поліпшуючої допоміжної інформації, яка пов'язана з одним або більше ієрархічними поліпшуючими рівнями. Базова допоміжна інформація може включати в себе базову незалежну допоміжну інформацію, яка стосується перших індивідуальних монауральних сигналів, які будуть декодуватися незалежно від інших монауральних сигналів. Кожний з одного або більше

25 ієрархічних поліпшуючих рівнів може включати в себе частину поліпшуючої допоміжної інформації, яка включає в себе параметри для поліпшення базового відтвореного представлення звуку, доступні з даних, включених у відповідні рівні і будь-які рівні нижче відповідного рівня.

Базова незалежна допоміжна інформація може вказувати, що перші індивідуальні монауральні сигнали представляють спрямований сигнал з напрямком падіння. Базова допоміжна інформація може додатково включати в себе базову залежну допоміжну інформацію, яка стосується других індивідуальних монауральних сигналів, які будуть декодуватися залежно від інших монауральних сигналів. Базова залежна допоміжна інформація може включати в себе

30 основані на векторах сигнали, які розподілені по напрямках у звуковому полі, причому розподіл по напрямках визначений за допомогою вектора. Компоненти вектора встановлені такими, що дорівнюють нулю, і не є частиною стисненого векторного представлення.

Компоненти базового стисненого представлення звуку можуть відповідати монауральним сигналам, які представляють або переважні звукові сигнали, або послідовності коефіцієнтів представлення HOA. Бітовий потік включає в себе корисні навантаження даних, відповідно

40 пов'язані з множиною ієрархічних рівнів. Поліпшуюча допоміжна інформація може включати в себе параметри, що стосуються щонайменше одного з перерахованого: просторове прогнозування, синтез спрямованих підсмугових сигналів і параметричне дублювання звукового оточення. Поліпшуюча допоміжна інформація може включати в себе інформацію, яка робить можливим прогнозування частин звуку або звукового поля, яких не вистачає, на основі

45 спрямованих сигналів. Може бути додатково визначено для кожного рівня, чи був відповідний рівень прийнятий коректно, і індекс рівня, що знаходиться безпосередньо нижче найбільш низького рівня, який не був прийнятий коректно.

Відповідно до іншого аспекту описана програма. Програма може бути адаптована для виконання на процесорі і для виконання деяких або всіх етапів способу, викладених у даному

50 документі, при її виконанні на обчислювальному пристрої.

Відповідно до ще одного аспекту описаний запам'ятовуючий носій. Запам'ятовуючий носій може містити програму, адаптовану для виконання на процесорі і для виконання деяких або всіх етапів способу, викладених у даному документі, при її виконанні на обчислювальному пристрої.

Твердження, зроблені відносно будь-якого зі згаданих вище аспектів або їх варіантів здійснення, також стосуються відповідних інших аспектів або їх варіантів здійснення, як зрозуміє фахівець в галузі техніки. Повторення цих тверджень для кожного аспекту або варіанта здійснення було опущено для стислості.

Способи і пристрої, що включають у себе переважні варіанти здійснення, викладені в даному документі, можуть використовуватися автономно або в сполученні з іншими способами і

60 системами, розкритими в цьому документі. Крім того, всі аспекти способів і пристроїв, викладені

в даному документі, можуть бути довільним чином об'єднані. Зокрема, ознаки пунктів формули винаходу можуть бути об'єднані одна з одною довільним чином.

Етапи способів і ознаки пристроїв можуть бути взаємозамінними різним чином. Зокрема, подробиці розкритого способу можуть бути реалізовані як пристрій, виконаний з можливістю виконувати деякі або всі етапи способу, і навпаки, як зрозуміє фахівець в галузі техніки.

КОРОТКИЙ ОПИС КРЕСЛЕНЬ

Винахід роз'яснений нижче ілюстративно з посиланням на прикладені креслення.

Фіг. 1 - блок-схема послідовності етапів, що ілюструє приклад способу багаторівневого кодування відповідно до варіантів здійснення розкриття.

Фіг. 2 - блок-схема, що схематично ілюструє приклад стадії кодера відповідно до варіантів здійснення розкриття.

Фіг. 3 - блок-схема послідовності етапів, що ілюструє приклад способу декодування стисненого представлення звуку або звукового поля, що було закодоване у множині ієрархічних рівнів, відповідно до варіантів здійснення розкриття.

Фіг. 4А і 4В - блок-схеми, що схематично ілюструють приклади стадії декодера, відповідно до варіантів здійснення розкриття.

Фіг. 5 - блок-схема, що схематично ілюструє приклад апаратної реалізації кодера відповідно до варіантів здійснення розкриття.

Фіг. 6 - блок-схема, що схематично ілюструє приклад апаратної реалізації декодера відповідно до варіантів здійснення розкриття.

ЗДІЙСНЕННЯ ВИНАХОДУ

Спочатку буде описане стиснене представлення звуку (або звукового поля) (далі для стислості називане стисненим представленням звуку), до якого застосовні способи і кодери/декодери відповідно до даного розкриття. У цілому повне стиснене представлення звуку (або звукового поля) (далі для стислості називане повним стисненим представленням звуку) може містити три наступні компоненти (наприклад, складатися з них): базове стиснене представлення звуку (або звукового поля) (далі для стислості називане базовим стисненим представленням звуку), базову допоміжну інформацію і поліпшуючу допоміжну інформацію.

Саме базове стиснене представлення звуку містить декілька компонентів (наприклад, складається з них) (наприклад, взаємодоповнюючих компонентів). Базове стиснене представлення звуку може брати до уваги визначено найбільший відсоток повного стисненого представлення звуку. Базове стиснене представлення звуку може складатися з монауральних транспортних сигналів, що представляють або переважні звукові сигнали, або послідовності коефіцієнтів початкового представлення НОА.

Базова допоміжна інформація потрібна для декодування базового стисненого представлення звуку і, як передбачається, має набагато менший розмір у порівнянні з базовим стисненим представленням звуку. Це може бути зроблено аж до її найбільшої частини незв'язних частин, кожна з яких визначає відновлення тільки одного конкретного компонента базового стисненого представлення звуку. Базова допоміжна інформація може містити першу частину, що може бути відома як незалежна базова допоміжна інформація, і другу частину, що може бути відома як додаткова базова допоміжна інформація.

І перша, і друга частини, незалежна базова допоміжна інформація і додаткова базова допоміжна інформація можуть визначати відновлення конкретних компонентів базового стисненого представлення звуку. Друга частина є факультативною і може бути опущена. У цьому випадку можна сказати, що стиснене представлення звуку містить першу частину (наприклад, базову допоміжну інформацію).

Перша частина (наприклад, базова допоміжна інформація) може містити допоміжну інформацію, яка описує індивідуальні (взаємодоповнюючі) компоненти базового стисненого представлення звуку, незалежно від інших (взаємодоповнюючих) компонентів. Зокрема, перша частина (наприклад, базова допоміжна інформація) може визначати декодування одного або більше з множини компонентів індивідуально, незалежно від інших компонентів. Таким чином, перша частина може згадуватися як незалежна базова допоміжна інформація.

Друга (факультативна) частина може містити допоміжну інформацію, також відому як додаткова базова допоміжна інформація, може описувати індивідуальні (взаємодоповнюючі) компоненти базового стисненого представлення звуку залежно від інших (взаємодоповнюючих) компонентів. Ця друга частина може також згадуватися як залежна базова допоміжна інформація. Зокрема, залежність може мати наступні властивості:

- залежна базова допоміжна інформація для кожного індивідуального (взаємодоповнюючого) компонента базового стисненого представлення звуку може досягати

свого найбільшого ступеня, коли інші визначені (взаємодоповнюючі) компоненти не містяться в базовому стисненому представленні звуку;

- у випадку, якщо додаткові визначені (взаємодоповнюючі) компоненти додані до базового стисненого представлення звуку, залежна базова допоміжна інформація для індивідуального (взаємодоповнюючого) компонента, що розглядається, може стати підмножиною початкової залежної базової допоміжної інформації, тим самим скорочуючи її розмір.

Поліпшуюча допоміжна інформація також є факультативною. Вона може використовуватися для поліпшення або розширення (наприклад, параметричного поліпшення або розширення) базового стисненого представлення звуку. Її розмір, як може також передбачатися, набагато менше, ніж у базового стисненого представлення звуку.

Таким чином, у варіантах здійснення стиснене представлення звуку може містити базове стиснене представлення звуку, що містить множину компонентів, базову допоміжну інформацію для декодування (наприклад, відновлення) базового стисненого представлення звуку до базового відтвореного представлення звуку або звукового поля і поліпшуючу допоміжну інформацію, яка включає в себе параметри для поліпшення або розширення (наприклад, параметричного поліпшення або розширення) базового відтвореного представлення звуку. Стиснене представлення звуку може також містити додаткову базову допоміжну інформацію для декодування (наприклад, відновлення) базового стисненого представлення звуку до базового відтвореного представлення звуку, що може включати в себе інформацію, яка визначає декодування одного або більше з множини компонентів залежно від відповідних інших компонентів.

Один приклад такого типу повного стисненого представлення звуку заданий за допомогою стисненого представлення Higher-Order Ambisonics (HOA) звукового поля (просторового звукового поля вищого порядку), як визначено за допомогою попередньої версії аудіостандарту MPEG-H 3D (посилання 1), глава 12 і додаток C.5. Таким чином, стиснене представлення звуку може відповідати стисненому представленню HOA звуку (або звукового поля).

Для цього прикладу базове стиснене представлення звукового поля (базове стиснене представлення звуку) може містити декілька компонентів (наприклад, може бути ідентифіковане за допомогою них). Компоненти можуть являти собою монауральні сигнали (наприклад, відповідати їм). Монауральні сигнали можуть являти собою квантовані монауральні сигнали. Монауральні сигнали можуть представляти або переважні звукові сигнали, або послідовності коефіцієнтів оточуючого компонента HOA звукового поля.

Базова допоміжна інформація може описувати, серед іншого, для кожного із цих монауральних сигналів, яким чином він додає просторовий внесок у звукове поле. Наприклад, базова допоміжна інформація може визначати переважний звуковий сигнал як чисто спрямований сигнал, що означає загальну плоску хвилю з деяким напрямком падіння. Як альтернатива базова допоміжна інформація може визначати монауральний сигнал як послідовність коефіцієнтів початкового представлення HOA, що має деякий індекс. Базова допоміжна інформація також може бути розділена на першу частину і другу частину, як зазначено вище.

Перша частина є допоміжною інформацією (наприклад, незалежною базовою допоміжною інформацією), яка стосується конкретних індивідуальних монауральних сигналів. Ця незалежна базова допоміжна інформація незалежна від існування інших монауральних сигналів. Така допоміжна інформація може, наприклад, визначати монауральний сигнал для представлення спрямованого сигналу (що, наприклад, означає загальну плоску хвилю) з деяким напрямком падіння. Як альтернатива монауральний сигнал може бути визначений як послідовність коефіцієнтів початкового представлення HOA, що має деякий індекс. Перша частина може згадуватися як незалежна базова допоміжна інформація. У цілому перша частина (наприклад, базова допоміжна інформація) може визначати декодування одного або більше з множини монауральних сигналів індивідуально, незалежно від інших монауральних сигналів.

Друга частина є допоміжною інформацією (наприклад, додатковою базовою допоміжною інформацією), яка стосується конкретних індивідуальних монауральних сигналів. Ця допоміжна інформація залежить від існування інших монауральних сигналів. Така допоміжна інформація може бути використана, наприклад, якщо монауральні сигнали визначені як основані на векторах сигнали (див., наприклад, посилання 1, розділ 12.4.2.4.4). Ці сигнали розподілені по напрямках у звуковому полі, причому розподіл по напрямках може бути визначений за допомогою вектора. У деякому режимі (див., наприклад, CodedVecLength=1) окремі компоненти цього вектора неявно встановлені такими, що дорівнюють нулю, і не є частиною стисненого векторного представлення. Цими компонентами є компоненти з індексами, що дорівнюють індексам послідовностей коефіцієнтів початкового представлення HOA і частини

базового стисненого представлення звуку. Це означає, що, якщо індивідуальні компоненти вектора закодовані, їх загальна кількість може залежати від базового стисненого представлення звуку. Зокрема, загальна кількість може залежати від того, які послідовності коефіцієнтів містить початкове представлення НОА.

5 Якщо послідовності коефіцієнтів початкового представлення НОА не містяться в базовому стисненому представленні звуку, залежна базова допоміжна інформація для кожного основаного на векторі сигналу складається із всіх векторних компонентів і має свій найбільший розмір. У випадку, якщо послідовності коефіцієнтів початкового представлення НОА з деякими індексами додаються до базового стисненого представлення звуку, векторні компоненти із цими індексами видаляються з допоміжної інформації для кожного основаного на векторі сигналу, тим самим скорочуючи розмір залежної базової допоміжної інформації для основаних на векторах сигналів.

10 Поліпшуюча допоміжна інформація (наприклад, поліпшуюча допоміжна інформація) може містити параметри, що стосуються (широкопasmогового) просторового прогнозування (див. посилання 1, розділ 12.4.2.4.3), і/або параметри, що стосуються синтезу спрямованих підсмугових сигналів і параметричного дублювання звукового оточення.

Параметри, що стосуються (широкопasmогового) просторового прогнозування, можуть використовуватися для (лінійного) прогнозування частин звукового поля, яких не вистачає, зі спрямованих сигналів.

20 Синтез спрямованих підсмугових сигналів і параметричне дублювання звукового оточення є інструментами стиснення, які були нещодавно введені в аудіостандарт MPEG-H 3D за допомогою поправки (див. посилання 2, розділ 1). Ці два інструменти дозволяють залежному від частоти параметричному прогнозуванню додаткових монауральних сигналів бути просторово розподіленим, щоб доповнювати просторово неповне або неповністю стиснене представлення НОА. Прогнозування може бути основане на послідовностях коефіцієнтів базового стисненого представлення звуку.

Важливо відзначити, що згаданий вище взаємодоповнюючий внесок у звукове поле представлений у стисненому представленні НОА не за допомогою додаткових квантованих сигналів, а за допомогою додаткової допоміжної інформації порівняно набагато меншого розміру. Отже, два згаданих інструменти кодування особливо придатні для стиснення представлень НОА на низьких швидкостях передачі даних.

30 Другий приклад стисненого представлення одного або більше монауральних сигналів зі згаданою вище структурою може містити закодовану спектральну інформацію для незв'язних частотних смуг аж до деякої верхньої частоти, що може розглядатися як базове стиснене представлення; базову допоміжну інформацію, яка визначає закодовану спектральну інформацію (наприклад, за допомогою кількості і ширини закодованих частотних смуг); і поліпшуючу допоміжну інформацію, яка містить параметри копіювання спектральної смуги (SBR) (наприклад, складається з них), які описують, як параметрично відтворити з базового стисненого представлення спектральну інформацію для смуг більш високої частоти, які не розглядаються в базовому стисненому представленні.

40 Дане розкриття пропонує спосіб багаторівневого кодування повного стисненого представлення звуку (або звукового поля), що має згадану вище структуру.

45 Стиснення може бути основане на кадрах у тому розумінні, що воно забезпечує стиснені представлення (у формі пакетів даних або еквівалентно корисного навантаження кадрів) для послідовних часових інтервалів. Часові інтервали можуть мати однакові або різні розміри. Ці пакети даних, як може передбачатися, містять прапор коректності, значення, що вказує їх розмір, а також фактичні дані стисненого представлення. Далі без навмисного обмеження буде передбачатися, що стиснення є основаним на кадрах. Крім того, якщо не зазначено інакше і без навмисного обмеження, буде зроблений фокус на обробці одного кадру, і тому індекс кадру

50 буде опущений.

Кожне корисне навантаження кадру повного стисненого представлення звуку (або звукового поля), що розглядається, як передбачається, містить J пакетів даних (або корисних навантажень кадрів), кожний для одного компонента базового стисненого представлення звуку, які позначені як $BSRC_j$, $j = 1, \dots, J$. Крім того, передбачається, що пакет містить незалежну базову допоміжну інформацію (базову допоміжну інформацію), позначену як BSI_I , яка визначає окремі компоненти $BSRC_j$ базового стисненого представлення звуку, незалежно від інших компонентів. Факультативно може додатково передбачатися, що пакет містить залежну базову допоміжну інформацію (додаткову базову допоміжну інформацію), позначену як BSI_D , яка

визначає окремі компоненти $BSRC_j$ базового стисненого представлення звуку залежно від інших компонентів.

Інформація, яка міститься у двох пакетах даних BSI_I і BSI_D , може бути факультативно згрупована в єдиний пакет даних BSI базової допоміжної інформації. Можна сказати, що єдиний пакет даних BSI містить, серед іншого, J частин, кожна з яких визначає один окремий компонент $BSRC_j$ базового стисненого представлення звуку. Можна сказати, що кожна із цих частин, у свою чергу, містить частину незалежної допоміжної інформації і факультативно частину залежної допоміжної інформації.

В остаточному підсумку, вона може включати в себе корисне навантаження поліпшуючої допоміжної інформації (поліпшуючої допоміжної інформації), позначене як ESI , з описом того, як поліпшити або розширити відтворений звук (або звукове поле) на основі повного базового стисненого представлення звуку.

Пропоноване рішення для багаторівневого кодування спрямоване на етапи, які вимагаються для забезпечення можливості як для частини стиснення, що включає в себе упакування пакетів даних для передачі, так і для частини прийому і відновлення. Кожна частина буде докладно описана далі.

Спочатку будуть описані стиснення і упакування (наприклад, для передачі). Зокрема, будуть описані компоненти і елементи повного стисненого представлення звуку (або звукового поля) у випадку багаторівневого кодування.

Фіг. 1 схематично ілюструє блок-схему послідовності етапів прикладу способу стиснення і упакування (наприклад, способу кодування або способу багаторівневого кодування стисненого представлення звуку або звукового поля). Присвоєння (наприклад, розподіл) індивідуальних корисних навантажень базовому рівню і (M-1) поліпшуючим рівням може бути досягнуте за допомогою пакувальника транспортних рівнів. Фіг. 2 схематично ілюструє блок-схему прикладу присвоєння/розподілу індивідуальних корисних навантажень.

Як зазначено вище, повне стиснене представлення 2100 звуку може стосуватися, наприклад, стисненого представлення НОА, що містить базове стиснене представлення звуку. Повне стиснене представлення 2100 звуку може містити множину компонентів (наприклад, монауральні сигнали) 2110-1, ..., 2110-J, незалежну базову допоміжну інформацію (базову допоміжну інформацію) 2120, факультативну поліпшуючу допоміжну інформацію (поліпшуючу допоміжну інформацію) 2140 і факультативну залежну базову допоміжну інформацію (додаткову базову допоміжну інформацію) 2130. Базова допоміжна інформація 2120 може бути інформацією для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля. Базова допоміжна інформація 2120 може включати в себе інформацію, яка визначає декодування одного або більше компонентів (наприклад, монауральних сигналів) індивідуально, незалежно від інших компонентів. Поліпшуюча допоміжна інформація 2140 може включати в себе параметри для поліпшення (наприклад, розширення) базового відтвореного представлення звуку. Додаткова базова допоміжна інформація 2130 може бути (додатковою) інформацією для декодування базового стисненого представлення звуку в базове відтворене представлення звуку і може включати в себе інформацію, яка визначає декодування одного або більше з множини компонентів залежно від відповідних інших компонентів.

Фіг. 2 ілюструє основне допущення, у якому існує множина ієрархічних рівнів, що включає у себе один базовий рівень (основний рівень) і один або більше (ієрархічних) поліпшуючих рівнів. Наприклад, може бути всього M рівнів, тобто один базовий рівень і M-1 поліпшуючих рівнів. Множина ієрархічних рівнів має послідовно збільшуваний індекс рівня. Найнижче значення індексу рівня (наприклад, індекс 1 рівня) відповідає базовому рівню. Далі мається на увазі, що рівні впорядковані від базового рівня, через поліпшуючі рівні, аж до повного найбільш високого поліпшуючого рівня (тобто повного найбільш високого рівня).

Запропонований спосіб може бути виконаний на основі кадру (тобто покадрово). Зокрема, стиснене представлення 2100 звуку може бути стиснене для послідовних часових інтервалів, наприклад часових інтервалів однакового розміру. Кожен часовий інтервал може відповідати кадру. Описані нижче етапи можуть бути виконані для кожного послідовного часового інтервалу (наприклад, кадру).

На етапі S1010 на фіг. 1 множина компонентів 2110 підрозділяється на множину груп компонентів. Кожна з множини груп потім присвоюється (наприклад, додається або розподіляється) відповідному одному з множини ієрархічних рівнів. При цьому кількість груп відповідає кількості рівнів. Наприклад, кількість груп може дорівнювати кількості рівнів, щоб була одна група компонентів для кожного рівня. Як зазначено вище, множина рівнів може

включати в себе базовий рівень і один або більше (наприклад, $M-1$) ієрархічних поліпшуючих рівнів.

Інакше кажучи, базове стиснене представлення звуку підрозділене на частини, які будуть присвоєні окремим рівням. Без втрати спільності утворювання може бути описане за допомогою $M+1$ чисел J_m , $m = 0, \dots, M$, де $J_0 = 1$ і $J_M = J + 1$, у результаті чого компоненти $BSRC_j$ присвоюються m -ому рівню для $J_{m-1} \leq j < J_m$.

На етапі S1020 групи компонентів присвоюються своїм відповідним рівням. На етапі S1030 базова допоміжна інформація 2120 додається (наприклад, розподіляється) до базового рівня (тобто до найбільш низького з множини ієрархічних рівнів).

Таким чином, внаслідок її невеликого розміру запропоновано включати повну базову допоміжну інформацію (базову допоміжну інформацію і факультативну додаткову базову допоміжну інформацію) у базовий рівень, щоб уникнути її непотрібної фрагментації.

Якщо стиснене представлення звуку, що розглядається, містить залежну базову допоміжну інформацію (додаткову базову допоміжну інформацію), спосіб додатково може включати (не показано на фіг. 1) декомпозицію додаткової базової допоміжної інформації на множини частин 2130-1, ..., 2130- M додаткової базової допоміжної інформації. Частини додаткової базової допоміжної інформації потім можуть бути додані (наприклад, розподілені) до базового рівня. Інакше кажучи, частини додаткової базової допоміжної інформації можуть бути включені в базовий рівень. Кожна частина додаткової базової допоміжної інформації може бути пов'язана з відповідним рівнем і може включати в себе інформацію, яка визначає декодування одного або більше компонентів, присвоєних відповідному рівню, залежно від інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня.

Таким чином, у той час як незалежна базова допоміжна інформація BSI_I (базова допоміжна інформація) 2120 залишається без змін для присвоєння, залежна базова допоміжна інформація повинна бути оброблена спеціально для багаторівневого кодування, щоб дозволити правильне декодування на стороні приймача, з одного боку, і скоротити розмір залежної базової допоміжної інформації для передачі, з іншої боку. Запропоновано виконати декомпозицію залежної базової допоміжної інформації на M частин, позначених як $BSI_{D,m}$, $m = 1, \dots, M$, де m -та частина містить залежну базову допоміжну інформацію для кожного з компонентів $BSRC_j$, $J_{m-1} \leq j < J_m$ базового стисненого представлення звуку, присвоєного m -ому рівню, у припущенні, що факультативна залежна базова допоміжна інформація існує для стисненого представлення звуку, що розглядається. У випадку, якщо відповідна залежна допоміжна інформація не існує, для стисненого представлення звуку частин $BSI_{D,m}$ може передбачатися порожньою. Кожна частина залежної базової допоміжної інформації $BSI_{D,m}$ може залежати від всіх компонентів $BSRC_j$, $1 \leq j < J_m$, що містяться на всіх рівнях аж до m -го (тобто містяться на всіх рівнях $j = 1, \dots, m$).

Якщо пакет BSI_I незалежної базової допоміжної інформації має нехтувано малий розмір, розумно утримувати його як ціле і додавати (присвоювати) його до базового рівня. Факультативно подібна декомпозиція, як для залежної базової допоміжної інформації, також може бути виконана для незалежної базової допоміжної інформації, забезпечуючи пакети $BSI_{I,m}$, $m = 1, \dots, M$. Це корисно для скорочення розміру базового рівня за допомогою додавання (присвоєння) частин незалежної базової допоміжної інформації до рівнів з відповідними компонентами базового стисненого представлення звуку.

На етапі S1040 може бути визначена множина частин 2140-1, ..., 2140- M поліпшуючої допоміжної інформації. Кожна частина поліпшуючої допоміжної інформації може включати в себе параметри для поліпшення (наприклад, розширення) відтвореного представлення звуку, доступні з даних, включених у відповідний рівень і будь-які рівні нижче відповідного рівня.

Причина виконання цього етапу полягає в тому, що у випадку багаторівневого кодування важливо реалізувати, щоб поліпшуюча допоміжна інформація обчислювалася для кожного додаткового рівня, оскільки передбачається поліпшити попередній відновлений звук (або звукове поле), що, однак, залежить від доступних рівнів для відновлення. Зокрема, попередній відновлений звук (або звукове поле) для даного найбільш високого декодованого рівня (найбільш високого застосовного рівня) залежить від компонентів, включених у найбільш високий декодований рівень і будь-які рівні нижче найбільш високого декодованого рівня. Отже, стиснення повинно забезпечити M індивідуальних пакетів даних поліпшуючої допоміжної інформації (частин поліпшуючої допоміжної інформації), позначених як ESI_m , $m = 1, \dots, M$, де

поліпшуюча допоміжна інформація в m -ому пакеті даних ESI_m обчислюється, щоб поліпшити представлення звуку (або звукового поля), одержане із всіх даних, що містяться на базовому рівні і поліпшуючих рівнях з індексами нижче m (наприклад, всіх даних, що містяться на m -ому рівні і будь-яких рівнях нижче m -го рівня).

5 На етапі S1050 множина частин 2140-1,..., 2140-M поліпшуючої допоміжної інформації присвоюється (наприклад, додана або розподіляється) множині рівнів. Кожна з множини частин поліпшуючої допоміжної інформації присвоюється відповідному одному з множини рівнів. Наприклад, кожний з множини рівнів включає в себе відповідну частину поліпшуючої допоміжної інформації.

10 Присвоєння базової і/або поліпшуючої допоміжної інформації відповідним рівням може бути зазначене в інформації конфігурації, яка формується за допомогою способу кодування. Інакше кажучи, відповідність між базовою і/або поліпшуючою допоміжною інформацією і відповідними рівнями може бути зазначена в інформації конфігурації. Крім того, інформація конфігурації може вказувати для кожного рівня компоненти базового стисненого представлення звуку, які присвоєні (наприклад, включені) цьому рівню. Частини додаткової базової допоміжної інформації, включені в базовий рівень, все ж таки можуть відповідати рівням, що відрізняються від базового рівня.

Пілбиваючи підсумок, на стадії стиснення забезпечується пакет даних кадру, позначений як FRAME що має наступний склад:

$$20 \quad \text{FRAME} = [\text{BSRC}_1 \quad \dots \quad \text{BSRC}_J \quad \text{BSI}_1 \quad \text{BSI}_{D,1} \quad \dots \quad \text{BSI}_{D,M} \quad \text{ESI}_1 \quad \dots \quad \text{ESI}_M] \quad (1)$$

Крім того, пакети $\text{BSI}_{I,i}$ і $\text{BSI}_{D,m}$ для $m = 1, \dots, M$ могли б бути об'єднані в єдиний пакет BSI , у цьому випадку пакет даних кадру, позначений як FRAME, буде мати наступний склад:

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

25 Порядок проходження індивідуальних корисних навантажень з пакетом даних кадру в загальному випадку може бути довільним.

Індивідуальні пакети даних потім можуть бути згруповані в корисних навантаженнях, які визначені як спеціальні пакети даних, які містять прапор коректності, значення, що вказує їх розмір, а також фактичні стиснені дані представлення. Використання корисних навантажень дозволяє просте демультимплексування на стороні приймача, пропонуючи перевагу можливості відкидати неактуальні корисні навантаження без необхідності їх аналізу. Одне можливе угруповання задане як

- присвоєння (наприклад, розподіл) кожного BSRC_j пакета $j = 1, \dots, J$ індивідуальному корисному навантаженню, позначеному як BP_j ;

35 - присвоєння (наприклад, розподіл) m -го пакета даних поліпшуючої допоміжної інформації ESI_m і m -го пакета даних $\text{BSI}_{D,m}$ залежної допоміжної інформації одному поліпшуючому корисному навантаженню, позначеному як EP_m , $m = 1, \dots, M$;

- присвоєння пакета незалежної базової допоміжної інформації BSI_I окремому корисному навантаженню допоміжної інформації, позначеному як BSIP .

Факультативно, якщо розмір незалежної базової допоміжної інформації великий, кожен m -ий з її компонентів, $\text{BSI}_{I,m}$, $m = 1, \dots, M$, може бути присвоєний (наприклад, розподілений) поліпшуючому корисному навантаженню EP_m . У цьому випадку корисне навантаження BSIP допоміжної інформації є порожнім і може бути проігнороване.

Інша факультативна можливість полягає в тому, щоб присвоїти всі залежні пакети даних $\text{BSI}_{D,m}$ базової допоміжної інформації корисному навантаженню BSIP допоміжної інформації, що є розумним, якщо розмір залежної базової допоміжної інформації є невеликим.

В остаточному підсумку може бути забезпечений пакет даних кадру, позначений як FRAME, що має наступний склад:

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

50 Порядок проходження індивідуальних корисних навантажень з пакетом даних кадру в загальному випадку може бути довільним.

Спосіб може додатково включати (не показано на фіг. 1) формування для кожного з множини рівнів пакета транспортного рівня (наприклад, пакета 2200 базового рівня і M-1 пакетів 2300-1,..., 2300-(M-1)) поліпшуючого рівня, що включають у себе дані відповідного рівня (наприклад, компоненти, базову допоміжну інформацію і поліпшуючу допоміжну інформацію для

базового рівня або компоненти і поліпшуючу допоміжну інформацію для одного або більше поліпшуваних рівнів).

Пакети транспортного рівня для різних рівнів можуть мати різні пріоритети передачі. Таким чином, спосіб може додатково включати (не показано на фіг. 1) формування транспортного потоку для передачі даних множини рівнів, причому базовий рівень має найбільш високий пріоритет передачі, і ієрархічні поліпшувачі рівні мають спадні пріоритети передачі. При цьому більш високий пріоритет передачі може відповідати більшому ступеню захисту від помилок, і навпаки.

Якщо етапи не вимагають деяких інших етапів як попередні умови, згадані вище етапи можуть виконуватися в будь-якому порядку, і передбачається, що ілюстративний порядок, показаний на фіг. 1, не має обмежувального характеру.

Фіг. 3 ілюструє спосіб декодування стисненого представлення звуку або звукового поля для декодування або відновлення. Приклади відповідної стадії прийому і відновлення схематично проілюстровані на блок-схемах на фіг. 4A і 4B.

Як випливає з попереднього опису, стиснене представлення звуку може бути закодоване в множині ієрархічних рівнів. Множина рівнів може мати присвоєні їм (наприклад, може включати в себе) компоненти базового стисненого представлення звуку, компоненти присвоюються відповідним рівням у відповідних групах компонентів. Базовий рівень може включати в себе базову допоміжну інформацію для декодування базового стисненого представлення звуку. Кожен рівень може включати в себе одну зі згаданих вище частин поліпшувачої допоміжної інформації, яка включає в себе параметри для поліпшення базового відтвореного представлення звуку, доступних з даних, включених у відповідний рівень і будь-які рівні нижче відповідного рівня.

Запропонований спосіб може бути виконаний на основі кадрів (тобто покaдрово). Зокрема, відновлене представлення звуку або звукового поля може бути сформоване для послідовних часових інтервалів, наприклад часових інтервалів однакового розміру. Часові інтервали можуть бути, наприклад, кадрами. Описані нижче етапи можуть бути виконані для кожних послідовних часових інтервалів (наприклад, кадрів).

На етапі S3010 приймаються корисні навантаження даних (наприклад, пакети транспортного рівня), що відповідають множині рівнів. Корисні навантаження даних можуть бути прийняті як частина бітового потоку, що містить стиснене представлення НОА звуку або звукового поля, представлення відповідає множині ієрархічних рівнів. Ієрархічні рівні включають в себе базовий рівень і один або більше ієрархічних поліпшуваних рівнів. Множина рівнів має присвоєні їм компоненти базового стисненого представлення звуку або звукового поля. Компоненти присвоєні відповідним рівням у відповідних групах компонентів.

Пакети індивідуальних рівнів можуть бути мультиплексовані для забезпечення прийнятого пакета кадру повного стисненого представлення звуку. Прийнятий пакет кадру може бути позначений

$$\left[BSI_1 \ BSI_{D,1} \ \dots \ BSI_{D,M} \ ESI_1 \ BSRC_1 \ \dots \ BSRC_{(J_1)-1} \ \dots \ ESI_M \ BSRC_{J_{(M-1)}} \ \dots \ BSRC_J \right]. \quad (4)$$

В альтернативному випадку пакети BSI_1 і $BSI_{D,m}$ для $m = 1, \dots, M$ об'єднані в єдиний пакет BSI , пакети індивідуальних рівнів можуть бути мультиплексовані для забезпечення прийнятого пакета $\left[BSI \ ESI_1 \ BSRC_1 \ \dots \ BSRC_{(J_1)-1} \ \dots \ ESI_M \ BSRC_{J_{(M-1)}} \ \dots \ BSRC_J \right]$. (5)

Прийнятий пакет кадру може бути заданий як

$$\text{FRAME} = \left[\overline{BP}_1 \ \dots \ \overline{BP}_J \ \overline{BSIP} \ \overline{EP}_1 \ \dots \ \overline{EP}_M \right]. \quad (6)$$

Прийнятий пакет кадру потім може бути переданий на декомпресор або декодер 4100. Якщо передача індивідуального рівня була безпомилковою, прапор коректності щонайменше частини вміщеного корисного навантаження \overline{EP}_m поліпшувачої допоміжної інформації (наприклад, відповідної частини поліпшувачої допоміжної інформації) установлений таким, що дорівнює "істинному". У випадку помилки внаслідок передачі індивідуального рівня прапор коректності щонайменше в корисному навантаженні поліпшувачої допоміжної інформації на цьому рівні встановлений таким, що дорівнює "хибному". Отже, коректність пакета рівня може бути визначена на основі коректності вміщеного корисного навантаження поліпшувачої допоміжної інформації (наприклад, на основі його прапора коректності).

У декомпресорі 4100 прийнятий пакет кадру може бути демультимплексований. Із цією метою може використовуватися інформація розміру кожного корисного навантаження, щоб уникнути непотрібного аналізу даних індивідуальних корисних навантажень.

На етапі S3020 перший індекс рівня, що вказує найбільш високий рівень (наприклад, найбільш високий застосовний рівень або найбільш високий декодований рівень), визначається з множини рівнів для використання для декодування базового стисненого представлення звуку в базове відтворене представлення звуку або звукового поля.

Крім того, на етапі S3020 може бути вибране значення (наприклад, індекс рівня) N_B найбільш високого рівня (найбільш високого застосовного рівня), що буде використовуватися для відновлення базового представлення звуку. Найбільш високий поліпшуючий рівень, що буде фактично використовуватися для відновлення базового представлення звуку, заданий як $N_B - 1$. Оскільки кожен рівень містить точно одне корисне навантаження поліпшуючої допоміжної інформації (частину поліпшуючої допоміжної інформації), можна визначити на основі корисного навантаження поліпшуючої допоміжної інформації, чи є коректним вміщуючий рівень (наприклад, був коректно прийнятий). Отже, вибір може бути досягнутий з використанням всіх корисних навантажень поліпшуючої допоміжної інформації $ESI_m, m = 1, \dots, M$ (або, відповідно, $EP_m, m = 1, \dots, M$).

На етапі S3030 одержується базове відтворене представлення звуку. Базове відтворене представлення звуку може бути одержане з компонентів, присвоєних найбільш високому застосовному рівню, зазначеному першим індексом рівня, і будь-яким рівням нижче цього найбільш високого застосовного рівня з використанням базової допоміжної інформації (або в цілому з використанням базової допоміжної інформації).

Корисні навантаження компонентів $BSRC_1, \dots, BSRC_J$ базового стисненого представлення звуку можуть бути забезпечені попарно з усіма корисними навантаженнями базової допоміжної інформації (наприклад, BSI або BSI_I і $BSI_{D,m}, m = 1, \dots, M$) і значенням N_B процесору 4200 відновлення базового представлення. Процесор 4200 відновлення базового представлення (проілюстрований на фіг. 4A і 4B) відтворює базове представлення звуку (або звукового поля) з використанням тільки тих компонентів базового стисненого представлення звуку, які містяться на найбільш низьких N_B рівнях, які являють собою базовий рівень і $N_B - 1$ поліпшуючих рівнів (тобто рівні аж до рівня, зазначеного першим індексом рівня). Як альтернатива процесору 4200 відновлення базового представлення можуть бути забезпечені тільки корисні навантаження компонентів базового стисненого представлення звуку, що містяться на найбільш низьких N_B рівнях разом з відповідними корисними навантаженнями базової допоміжної інформації.

Необхідна інформація про те, які компоненти базового стисненого представлення звуку (або звукового поля) містяться на індивідуальних рівнях, передбачається відомою декомпресору 4100 з пакета даних з інформацією конфігурації, яка передбачається відправленою і прийнятою перед пакетами даних кадру.

Щоб забезпечити пакети даних $BSI_{D,m}, m = 1, \dots, N_B$ залежної допоміжної інформації і пакет даних ESI_{N_E} поліпшуючої допоміжної інформації, всі поліпшуючі корисні навантаження можуть введені в частковий аналізатор 4400 (див. фіг. 4B) декомпресора 4100 разом зі значенням N_E і значенням N_B . Аналізатор може відкинути всі корисні навантаження і пакети даних, які не будуть використовуватися для фактичного відновлення. Якщо значення N_E дорівнює нулю, то може передбачатися, що всі пакети даних поліпшуючої допоміжної інформації є порожніми.

Якщо базовий рівень включає в себе щонайменше одне залежне корисне навантаження базової допоміжної інформації (частину додаткової базової допоміжної інформації), що відповідає відповідному рівню, декодування кожного індивідуального корисного навантаження залежної базової допоміжної інформації (наприклад, $BSI_{D,m}, m = 1, \dots, N_B$ (частина додаткової базової допоміжної інформації)) може включати в себе (i) декодування частини додаткової базової допоміжної інформації за допомогою посилення на компоненти, присвоєні їй відповідному рівню і будь-яким рівням нижче відповідного рівня (попереднє декодування), і (ii) корекцію частини додаткової базової допоміжної інформації за допомогою посилення на компоненти, присвоєні найбільш високому застосовному рівню і будь-яким рівням між найбільш високим застосовним рівнем і відповідним рівнем (корекція). При цьому додаткова базова допоміжна інформація, що відповідає відповідному рівню, включає в себе інформацію, яка визначає декодування одного або більше компонентів з компонентів, присвоєних відповідному

рівню, залежно від інших компонентів, присвоєних відповідному рівню і будь-яким рівням нижче відповідного рівня.

Потім базове відтворене представлення звуку може бути одержане (наприклад, сформоване) з компонентів, присвоєних найбільш високому застосовному рівню і будь-яким рівням нижче найбільш високого застосовного рівня, з використанням базової допоміжної інформації і скоректованих частин додаткової базової допоміжної інформації, одержаних із частин додаткової базової допоміжної інформації, що відповідає рівням аж до найбільш високого застосовного рівня.

Зокрема, попереднє декодування кожного корисного навантаження $BSI_{D,m}$, $m = 1, \dots, N_B$, може включати в себе використання його залежності від перших $J_m - 1$ компонентів $BSRC_1, \dots, BSRC_{(J_m) - 1}$ базового стисненого представлення звуку, які містяться на перших m рівнях, що передбачалося на стадії кодування.

Послідовна корекція кожного корисного навантаження $BSI_{D,m}$, $m = 1, \dots, N_B$, може включати в себе прийняття до уваги, що базовий компонент звуку нарешті відтворений з перших $J_{N_B} - 1$ компонентів $BSRC_1, \dots, BSRC_{(J_{N_B}) - 1}$ базового стисненого представлення звуку, які містяться на перших $N_B > m$ рівнях, що є більшою кількістю компонентів, ніж передбачалося для попереднього декодування.

Отже, корекція може бути досягнута за допомогою відкидання неадекватної інформації, що можливе внаслідок початково прийнятої властивості залежної базової допоміжної інформації, що полягає в тому, що, якщо деякі взаємодоповнюючі компоненти додаються до базового стисненого представлення звуку, залежна базова допоміжна інформація для кожного індивідуального (взаємодоповнюючого) компонента стає підмножиною початкової.

На етапі S3040 може бути визначений другий індекс рівня. Другий індекс рівня може вказувати частину (частини) поліпшуючої допоміжної інформації, яка повинна використовуватися для поліпшення (наприклад, розширення) базового відтвореного представлення звуку.

На доповнення до першого індексу рівня може бути визначений індекс N_E (другий індекс рівня) корисного навантаження поліпшуючої допоміжної інформації (частини другої поліпшуючої інформації) для використання для відновлення. Другий індекс N_E рівня може завжди або дорівнювати першому індексу N_B рівня, або дорівнювати нулю. Поліпшення може бути досягнуте або завжди відповідно до базового представлення звуку, одержаного з найбільш високого застосовного рівня, або ніколи.

На етапі S3050 відтворене представлення звуку або звукового поля одержується (наприклад, формується) з базового відтвореного представлення звуку з посиланням на другий індекс рівня.

Таким чином, відтворене представлення звуку одержується за допомогою (параметричного) поліпшення або розширення базового відтвореного представлення звуку, наприклад за допомогою використання поліпшуючої допоміжної інформації (частини поліпшуючої допоміжної інформації), зазначеної другим індексом рівня. Як зазначено далі, другий індекс рівня може вказувати на те, щоб взагалі не використовувати яку-небудь поліпшуючу допоміжну інформацію на даній стадії. Тоді відтворене представлення звуку буде відповідати базовому відтвореному представленню звуку.

Із цією метою відтворене базове представлення звуку разом з усіма корисними навантаженнями ESI_1, \dots, ESI_M поліпшуючої допоміжної інформації, корисними навантаженнями базової допоміжної інформації (наприклад, BSI або BSI_i і $BSI_{D,m}$, $m = 1, \dots, M$) і значенням N_E забезпечуються процесору 4300 відновлення розширеного представлення (проілюстрованому на фіг. 4A і 4B), який обчислює остаточне розширене представлення 2100' звуку (або звукового поля) з використанням тільки корисного навантаження ESI_{N_E} поліпшуючої допоміжної інформації, і відкидаючи всі інші корисні навантаження поліпшуючої допоміжної інформації. Як альтернатива процесору 4300 відновлення поліпшуючого представлення може бути забезпечене тільки корисне навантаження ESI_{N_E} поліпшуючої допоміжної інформації замість всіх корисних навантажень поліпшуючої допоміжної інформації. Якщо значення N_E дорівнює нулю, всі корисні навантаження поліпшуючої допоміжної інформації відкидаються (або як

альтернатива корисне навантаження поліпшуючої допоміжної інформації не забезпечується), і відтворене фінальне розширене представлення звуку дорівнює відтвореному основному представленню звуку. Корисне навантаження ESI_{N_E} поліпшуючої допоміжної інформації може бути одержане за допомогою часткового аналізатора 4400.

Фіг. 3 також у цілому ілюструє декодування стисненого представлення НОА на основі базової допоміжної інформації, що пов'язана з базовим рівнем, і на основі поліпшуючої допоміжної інформації, що пов'язана з одним або більше ієрархічними поліпшуваними рівнями.

Якщо етапи не вимагають деяких інших етапів як попередні умови, згадані вище етапи можуть виконуватися в будь-якому порядку, і передбачається, що ілюстративний порядок, показаний на фіг. 3, не має обмежувального характеру.

Далі будуть описані подробиці вибору рівнів для відновлення (вибір першого і другого індексів рівнів) на етапах S3020 і S3040.

Визначення першого індексу рівня може включати в себе визначення для кожного рівня, чи був відповідний рівень прийнятий коректно. Визначення першого індексу рівня може додатково включати в себе визначення першого індексу рівня як індексу того рівня, який знаходиться безпосередньо нижче найбільш низького рівня, що не був коректно прийнятий. Чи був рівень прийнятий коректно, може бути визначено за допомогою оцінки, чи було коректно прийняте корисне навантаження поліпшуючої допоміжної інформації цього рівня. Це, у свою чергу, може бути виконане за допомогою оцінки прапорів коректності в корисних навантаженнях поліпшуючої допоміжної інформації.

Визначення другого індексу рівня в загальному випадку може включати в себе або визначення другого індексу рівня як такого, що дорівнює першому індексу рівня, або визначення значення індексу як другого індексу рівня (наприклад, значення 0 індексу), яке вказує, що не слід використовувати яку-небудь поліпшуючу допоміжну інформацію при одержанні відтвореного представлення звуку.

У випадку, якщо всі пакети даних кадру можуть бути відновлені незалежно один від одного, і номер N_B найбільш високого рівня (найбільш високого застосовного рівня) для фактичного використання для відновлення базового представлення звуку, і індекс N_E корисного навантаження поліпшуючої допоміжної інформації для використання для відновлення можуть бути встановлені такими, що дорівнюють найбільшому номеру L коректного корисного навантаження поліпшуючої допоміжної інформації, який сам може бути визначений за допомогою оцінки прапорів коректності в корисних навантаженнях поліпшуючої допоміжної інформації. Використовуючи знання розміру кожного корисного навантаження поліпшуючої допоміжної інформації, можна уникнути складного аналізу фактичних даних корисних навантажень для визначення їх коректності.

Таким чином, другий індекс рівня може бути визначений як такий, що дорівнює першому індексу рівня, якщо стиснені представлення звуку для послідовних часових інтервалів можуть бути декодовані незалежно. У цьому випадку відтворене базове представлення звуку може бути розширене на основі корисного навантаження поліпшуючої допоміжної інформації найбільш високого застосовного рівня.

У випадку, якщо використовується це диференціальне відновлення з міжкадровими залежностями, на доповнення потрібно розглядати рішення від попереднього кадру. Слід зазначити, що з диференціальним відновленням звичайно незалежні пакети даних кадру передаються з регулярними часовими інтервалами, щоб дозволити починати відновлення з тих моментів часу, коли визначення значень N_B і N_E стають незалежними від кадрів, і це виконується, як описано вище.

Для докладного роз'яснення запропонованого залежного від кадрів рішення, найбільший номер (наприклад, індекс рівня) коректного корисного навантаження поліпшуючої допоміжної інформації для k -го кадру позначений як $L(k)$, номер найбільш високого рівня (наприклад, індекс рівня) для вибору і використання для відновлення базового представлення звуку позначений як $N_B(k)$, і номер (наприклад, індекс рівня) корисного навантаження поліпшуючої допоміжної інформації для використання для відновлення позначений як $N_E(k)$.

Використовуючи ці позначення, номер найбільш високого рівня для використання для відновлення базового представлення звуку $N_B(k)$ може бути обчислений відповідно до

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

За допомогою вибору $N_B(k)$ не більше, ніж $N_B(k-1)$ і $L(k)$, забезпечується, що вся інформація, необхідна для диференціального відновлення базового представлення звуку, є доступною.

Таким чином, якщо стиснені представлення звуку для послідовних часових інтервалів (наприклад, кадрів) не можуть бути декодовані незалежно одне від одного, визначення першого індексу рівня може включати визначення для кожного рівня, чи був відповідний рівень прийнятий коректно, і визначення першого індексу рівня для даного часового інтервалу як меншого індексу з першого індексу рівня часового інтервалу, що передує даному часовому інтервалу, і індексу рівня, що знаходиться безпосередньо нижче найбільш низького рівня, що не був коректно прийнятий.

Номер $N_E(k)$ корисного навантаження поліпшуючої допоміжної інформації для вики $N_E(k) = \begin{cases} N_B(k) & \text{if } N_B(k) = N_B(k-1) \\ 0 & \text{else} \end{cases}$ визначений відповідно до . (8)

При цьому вибір 0 для $N_E(k)$ вказує, що відтворене базове представлення звуку не повинно поліпшуватися або розширюватися з використанням поліпшуючої допоміжної інформації.

Це означає, зокрема, що, за умови, що номер $N_B(k)$ найбільш високого рівня для використання для відновлення базового представлення звуку не змінюється, вибирається той же самий відповідний номер поліпшуючого рівня. Однак у випадку зміни $N_B(k)$ поліпшення забороняється за допомогою установа $N_E(k)$ таким, що дорівнює нулю. Внаслідок передбачуваного диференціального відновлення поліпшуючої допоміжної інформації її зміна відповідно до $N_B(k)$ неможлива, оскільки це вимагало б відновлення відповідного рівня поліпшуючої допоміжної інформації в попередньому кадрі, що, як передбачається, не було виконано.

Таким чином, якщо стиснені представлення звуку для послідовних часових інтервалів (наприклад, кадрів) не можуть бути декодовані незалежно одне від одного, визначення другого індексу рівня може включати визначення, чи дорівнює перший індекс рівня для даного часового інтервалу першому індексу рівня для попереднього часового інтервалу. Якщо перший індекс рівня для даного часового інтервалу дорівнює першому індексу рівня для попереднього часового інтервалу, другий індекс рівня для даного часового інтервалу може бути визначений (наприклад, вибраний) як такий, що дорівнює першому індексу рівня для даного часового інтервалу. З іншого боку, якщо перший індекс рівня для даного часового інтервалу не дорівнює першому індексу рівня для попереднього часового інтервалу, значення індексу може бути визначене (наприклад, вибране) як другий індекс рівня, який вказує, що не слід використовувати яку-небудь поліпшуючу допоміжну інформацію при одержанні відтвореного представлення звуку.

Як альтернатива, якщо при відновленні всі корисні навантаження поліпшуючої допоміжної інформації з номером аж до $N_E(k)$ відновлені паралельно, правило вибору в рівнянні (4) може бути $N_E(k) = N_B(k)$. (9)

Нарешті, слід зазначити, що для диференціального відновлення номер найбільш високого використовуваного рівня N_B може тільки збільшуватися в незалежних пакетах даних кадру, тоді як зменшення можливе в кожному кадрі.

Мається на увазі, що запропонований спосіб багаторівневого кодування стисненого представлення звуку може бути реалізований кодером для багаторівневого кодування стисненого представлення звуку. Такий кодер може містити відповідні блоки, виконані з можливістю виконувати відповідні описані вище етапи. Приклад такого кодера 5000 схематично проілюстрований на фіг. 5. Наприклад, такий кодер 5000 може містити блок 5010 підрозділення компонентів, виконаний з можливістю виконувати згаданий вище етап S1010, блок 5020 присвоєння компонентів, виконаний з можливістю виконувати згаданий вище етап S1020, блок 5030 присвоєння базової допоміжної інформації, виконаний з можливістю виконувати згаданий вище етап S1030, блок 5040 розбивки поліпшуючої допоміжної інформації, виконаний з можливістю виконувати згаданий вище етап S1040, і блок 5050 присвоєння поліпшуючої допоміжної інформації, виконаний з можливістю виконувати згаданий вище етап S1050. Далі мається на увазі, що відповідні блоки такого кодера можуть бути реалізовані за допомогою

процесора 5100 обчислювального пристрою, який виконаний з можливістю виконувати обробку, виконувану кожним зі згаданих відповідних блоків, тобто він виконаний з можливістю виконувати деякі або всі згадані вище етапи, а також будь-які додаткові етапи запропонованого методу кодування. Кодер або обчислювальний пристрій може додатково містити пам'ять 5200, до якої процесор 5100 може здійснювати доступ.

Далі мається на увазі, що запропонований спосіб декодування стисненого представлення звуку, яке закодоване в множині ієрархічних рівнів, може бути реалізований декодером для декодування стисненого представлення звуку, яке закодоване в множині ієрархічних рівнів. Такий декодер може містити відповідні блоки, виконані з можливістю виконувати відповідні описані вище етапи. Приклад такого декодера 6000 схематично проілюстрований на фіг. 6. Наприклад, такий декодер 6000 може містити блок 6010 прийому, виконаний з можливістю виконувати згаданий вище етап S3010, блок 6020 визначення першого індексу рівня, виконаний з можливістю виконувати згаданий вище етап S3020, блок 6030 базового відтворення, виконаний з можливістю виконувати згаданий вище етап S3030, блок 6040 визначення другого індексу рівня, виконаний з можливістю виконувати згаданий вище етап S3040, і блок 6050, виконаний з можливістю виконувати згаданий вище етап S3050. Далі мається на увазі, що відповідні блоки такого кодера можуть бути реалізовані за допомогою процесора 6100 обчислювального пристрою, який виконаний з можливістю виконувати обробку, виконувану кожним зі згаданих відповідних блоків, тобто він виконаний з можливістю виконувати деякі або всі згадані вище етапи, а також будь-які додаткові етапи запропонованого методу кодування. Кодер або обчислювальний пристрій може додатково містити пам'ять 6200, до якої процесор 6100 може здійснювати доступ.

Слід зазначити, що опис і креслення лише ілюструють принципи запропонованих способів і пристроїв. Таким чином, буде очевидно, що фахівці в галузі техніки зможуть створювати різні структури, які, хоча явно не описані і не показані в даному документі, реалізують принципи винаходу і включені в межі його суті і обсягу. Крім того, всі приклади, наведені в даному документі, переважно явно призначені лише для навчання, щоб допомогти читачеві в розумінні принципів запропонованих способів і пристроїв, і концепцій, внесених винахідниками в розвиток галузі техніки, і повинні бути витлумачені як такі, що не є обмеженнями для таких спеціальним чином наведених прикладів і умов. Крім того, передбачається, що всі твердження в даному документі, що викладають принципи, аспекти і варіанти здійснення винаходу, а також їх конкретні приклади, охоплюють його еквіваленти.

Способи і пристрій, описані в даному документі, можуть бути реалізовані як програмне забезпечення, програмно-апаратне забезпечення і/або апаратні засоби. Деякі компоненти, наприклад, можуть бути реалізовані як програмне забезпечення, що працює на процесорі цифрової обробки сигналів або мікропроцесорі. Інші компоненти, наприклад, можуть бути реалізовані як апаратні засоби і/або як спеціалізовані інтегральні схеми. Сигнали, що зустрічаються в описаних способах і пристрої, можуть бути збережені на носіях, таких як оперативний запам'ятовуючий пристрій або оптичний запам'ятовуючий носії. Вони можуть бути перенесені через мережі, такі як радіомережі, супутникові мережі, безпроводні мережі або провідні мережі, наприклад Інтернет.

Джерела інформації:

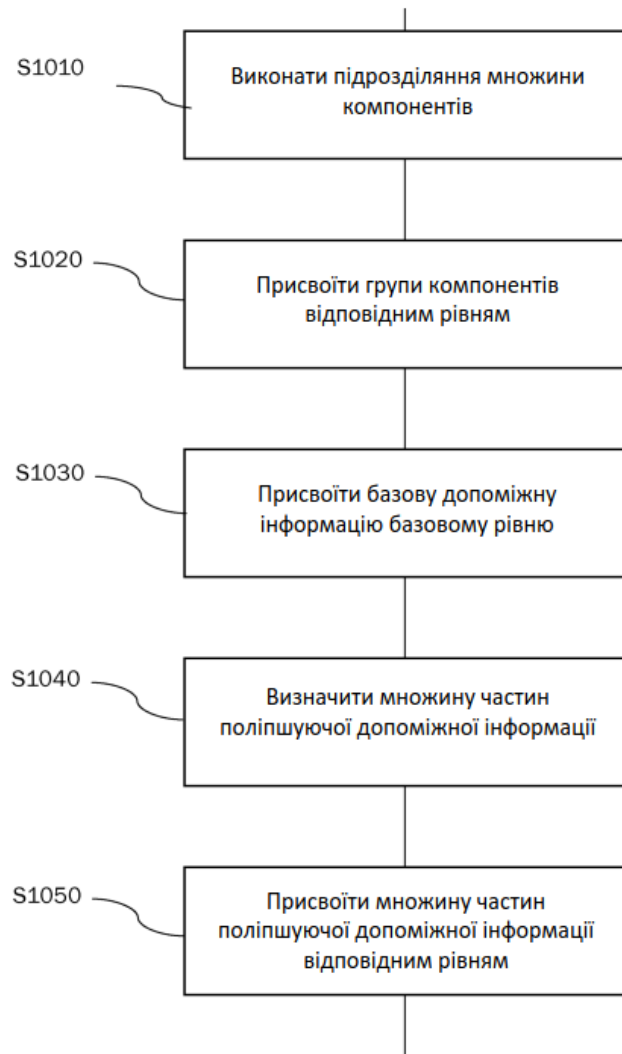
1. ISO/IEC JTC1/SC29/WG11 23008-3:2015(E). Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3: 3D audio, February 2015.
2. ISO/IEC JTC1/SC29/WG11 23008-3:2015/PDAM3. Information technology-High efficiency coding and media delivery in heterogeneous environments-Part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

ФОРМУЛА ВИНАХОДУ

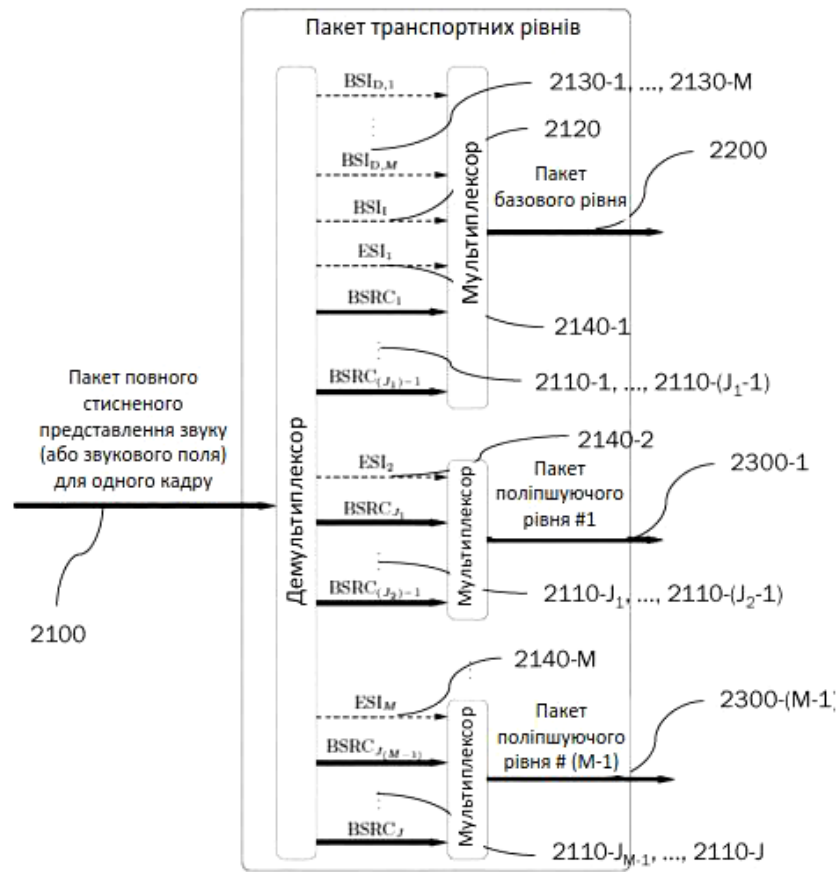
1. Спосіб декодування стисненого представлення звуку або звукового поля системи Амбісонік вищого порядку (HOA), що кодується в множині ієрархічних рівнів, використовуючи рівневе кодування, причому спосіб включає етапи, на яких:
 - приймають (S_{3010}) бітовий потік, що містить стиснене представлення HOA, що відповідає множині ієрархічних рівнів, які включають в себе базовий рівень (2200) і щонайменше два ієрархічні поліпшуючі рівні (2300), причому множина рівнів мають присвоєні їм компоненти базового стисненого представлення звуку або звукового поля, причому компоненти відповідають множині монауральних сигналів і присвоюються відповідним рівням у відповідних групах компонентів, і

- декодуєть стиснене представлення НОА на основі базової допоміжної інформації (2120), яка пов'язана з базовим рівнем (2200), і на основі поліпшуючої допоміжної інформації (2140), яка пов'язана зі щонайменше двома ієрархічними поліпшуваними рівнями (2300),
- 5 при цьому базова допоміжна інформація (2120) включає в себе базову незалежну допоміжну інформацію (2120), що стосується перших індивідуальних монауральних сигналів множини монауральних сигналів, які будуть декодовані незалежно від інших монауральних сигналів множини монауральних сигналів.
2. Спосіб за п. 1, в якому базова незалежна допоміжна інформація (2120) вказує, що перші індивідуальні монауральні сигнали представляють направлений сигнал з напрямком падіння.
- 10 3. Спосіб за будь-яким із пп. 1-2, в якому базова допоміжна інформація (2120) додатково включає в себе базову залежну допоміжну інформацію (2130), що стосується других індивідуальних монауральних сигналів множини монауральних сигналів, які будуть декодовані залежно від інших монауральних сигналів множини монауральних сигналів.
4. Спосіб за п. 3, в якому базова залежна допоміжна інформація (2130) включає в себе основані на векторах сигнали, які розподілені у напрямках в звуковому полі, причому розподіл у
- 15 напрямках визначений за допомогою вектора.
5. Спосіб за п. 4, в якому компоненти вектора встановлені такими, що дорівнюють нулю і не є частиною стисненого векторного представлення.
6. Спосіб за будь-яким із пп. 1-5, в якому монауральні сигнали представляють або переважні
- 20 звукові сигнали, або послідовності коефіцієнтів представлення НОА.
7. Спосіб за будь-яким із пп. 1-6, в якому бітовий потік включає в себе корисні навантаження даних, відповідно пов'язані з множиною ієрархічних рівнів.
8. Спосіб за будь-яким із пп. 1-7, в якому поліпшуюча допоміжна інформація (2140) включає в себе параметри, що стосуються щонайменше одного з перерахованого: просторове
- 25 прогнозування, синтез направлених підсмугових сигналів і параметричне дублювання звукового оточення.
9. Спосіб за будь-яким із пп. 1-8, в якому поліпшуюча допоміжна інформація (2140) включає в себе інформацію, яка забезпечує можливість прогнозування частин звуку, яких не вистачає, або звукового поля з направлених сигналів.
- 30 10. Спосіб за будь-яким із пп. 1-9, який додатково містить етапи, на яких: визначають для кожного рівня, чи був відповідний рівень прийнятий коректно; і визначають індекс рівня, що знаходиться безпосередньо нижче найбільш низького рівня, який не був прийнятий коректно.
11. Пристрій (6000) для декодування стисненого представлення звуку або звукового поля системи Амбісонік вищого порядку (НОА), який кодується в множині ієрархічних рівнів, використовуючи рівневе кодування, причому пристрій (6000) містить:
- 35 приймач (6010) для прийому бітового потоку, що містить стиснене представлення НОА, що відповідає множині ієрархічних рівнів, які включають в себе базовий рівень (2200) і щонайменше два ієрархічні поліпшувачі рівні (2300), причому множина рівнів мають присвоєні їм компоненти базового стисненого представлення звуку або звукового поля, причому компоненти
- 40 відповідають множині монауральних сигналів і присвоюються відповідним рівням у відповідних групах компонентів,
- декодер для декодування стисненого представлення НОА на основі базової допоміжної інформації (2120), яка пов'язана з базовим рівнем (2200), і на основі поліпшуючої допоміжної
- 45 інформації (2140), яка пов'язана зі щонайменше двома ієрархічними поліпшуваними рівнями (2300),
- при цьому базова допоміжна інформація (2120) включає в себе базову незалежну допоміжну інформацію (2120), що стосується перших індивідуальних монауральних сигналів множини монауральних сигналів, які будуть декодовані незалежно від інших монауральних сигналів
- 50 множини монауральних сигналів.
12. Пристрій (6000) за п. 11, в якому базова незалежна допоміжна інформація (2120) включає в себе визначення щонайменше монаурального сигналу для представлення направлено
- сигналу з напрямком падіння.
13. Пристрій (6000) за будь-яким із пп. 11-12, в якому базова допоміжна інформація (2120)
- 55 додатково включає в себе базову залежну допоміжну інформацію (2130), що стосується других індивідуальних монауральних сигналів множини монауральних сигналів, які будуть декодовані залежно від інших монауральних сигналів множини монауральних сигналів.
14. Пристрій (6000) за п. 13, в якому базова залежна допоміжна інформація (2130) включає в себе основані на векторах сигнали, які розподілені у напрямках в звуковому полі, причому
- 60 розподіл у напрямках визначений за допомогою вектора.

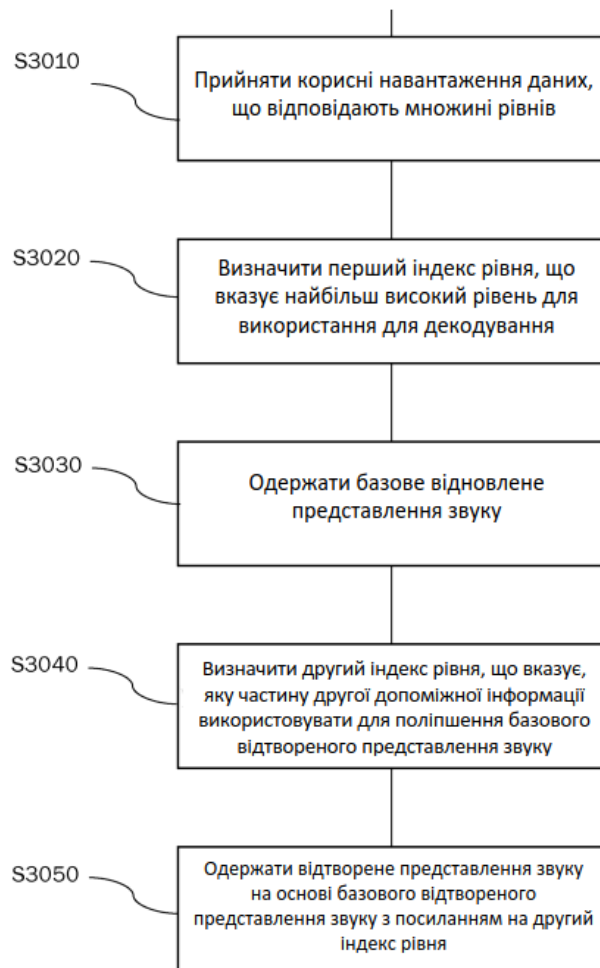
15. Пристрій (6000) за п. 14, в якому компоненти вектора встановлені такими, що дорівнюють нулю і не є частиною стисненого векторного представлення.



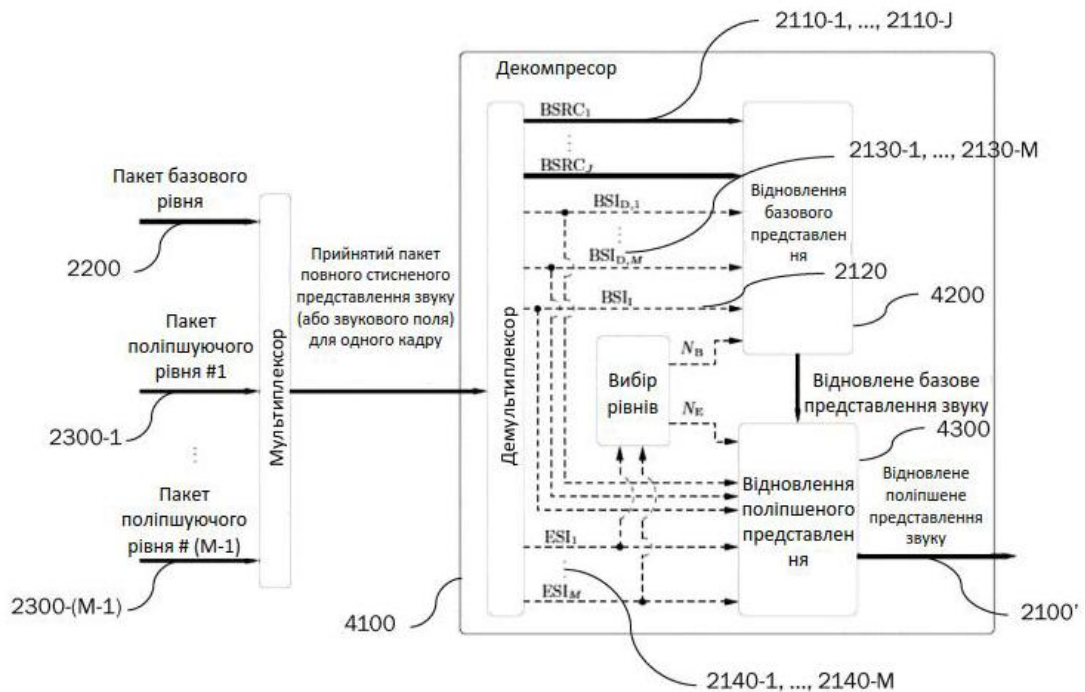
Фіг. 1



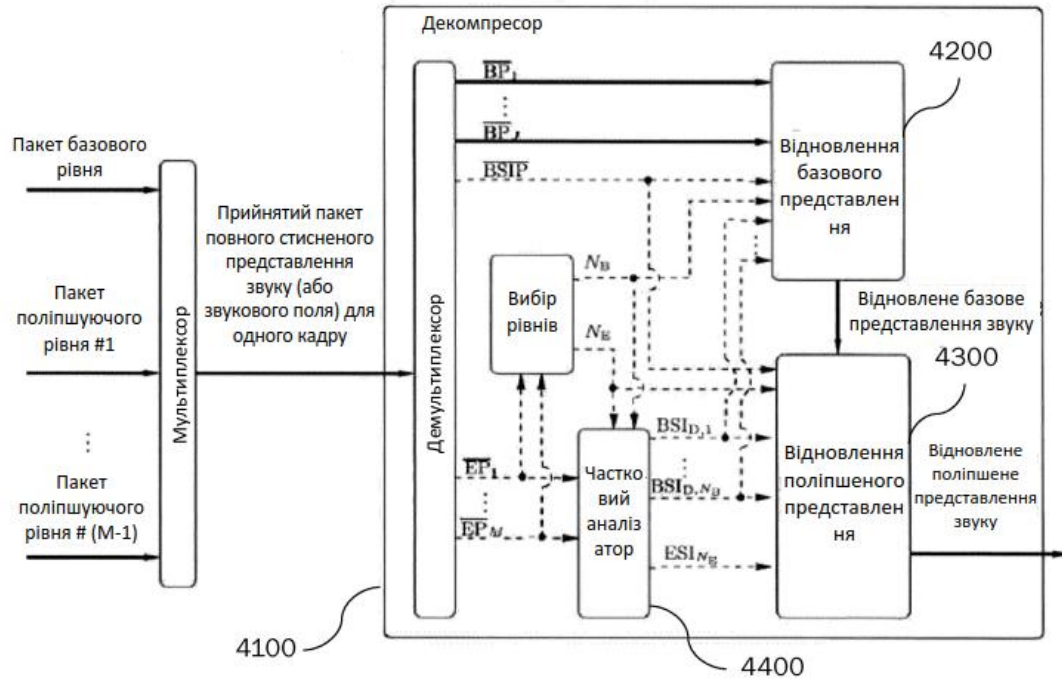
Фіг. 2



Фіг. 3



Фіг. 4A



Фіг. 4В

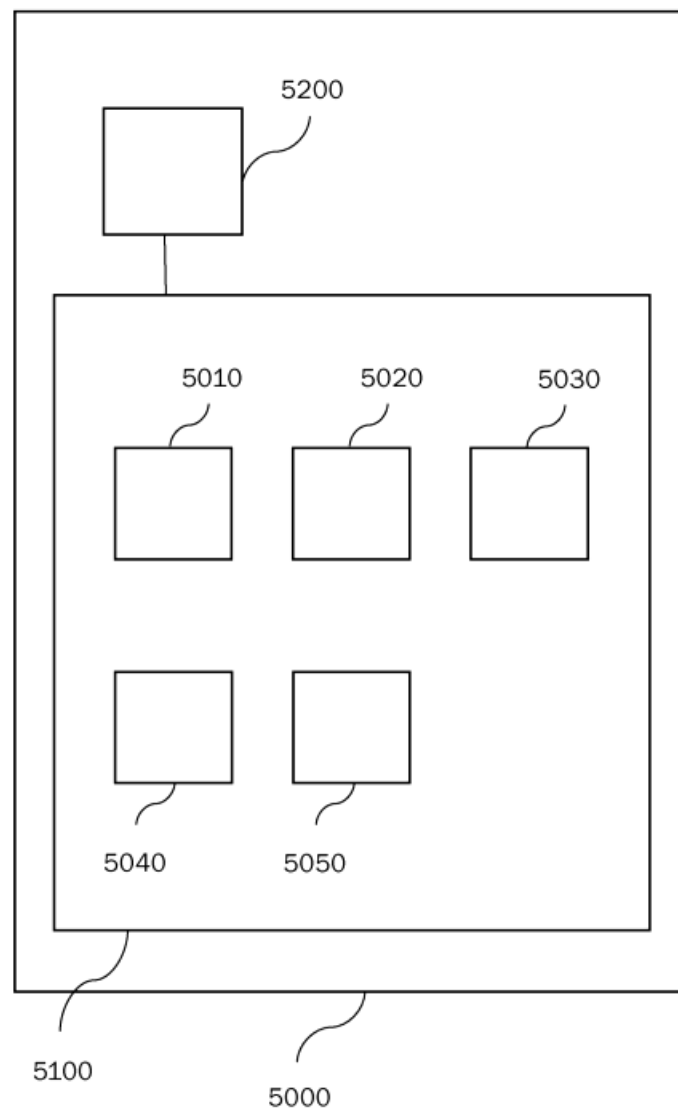
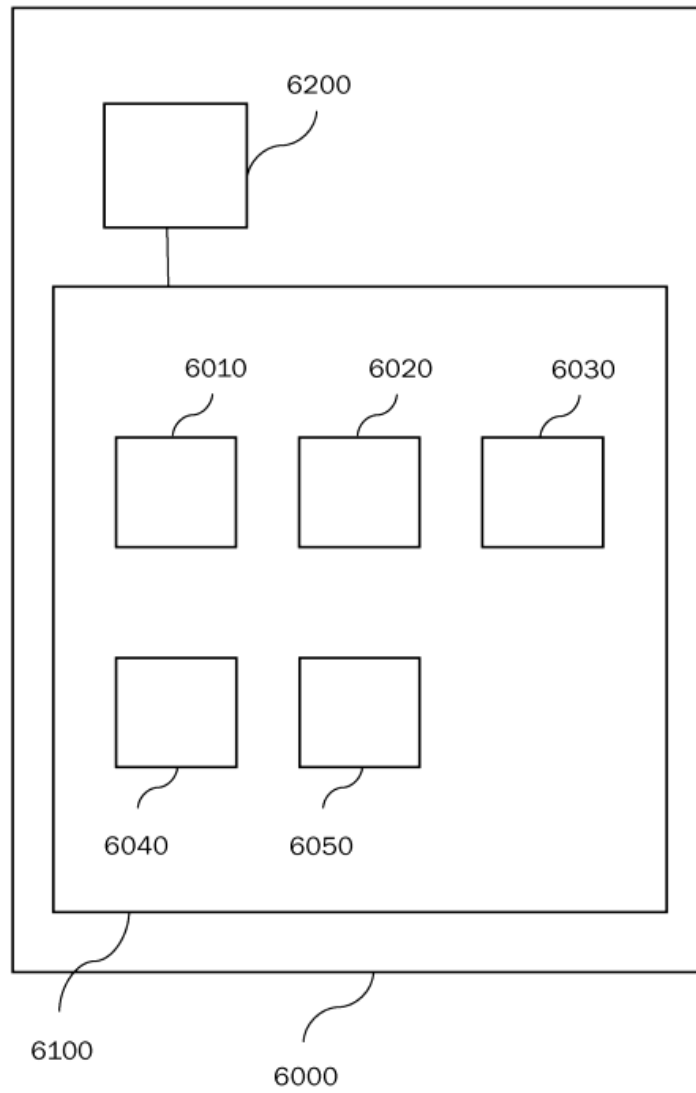


Fig. 5



Фіг. 6