(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0155983
(43) 공개일자 2024년10월29일

(51) 국제특허분류(Int. Cl.)
G10L 19/008 (2014.01) *G10L 19/16* (2013.01)
G10L 19/24 (2013.01) *H04S 7/00* (2006.01)
(52) CPC특허분류
G10L 19/008 (2020.08)
G10L 19/167 (2013.01)
(21) 출원번호 10-2024-7035074(분할)
(22) 출원일자(국제) 2019년04월09일
심사청구일자 2024년10월21일
(62) 원출원 특허 10-2020-7024701
원출원일자(국제) 2019년04월09일
심사청구일자 2022년04월05일
(85) 번역문제출일자 2024년10월21일
(86) 국제출원번호 PCT/EP2019/058955
(87) 국제공개번호 WO 2019/197404
국제공개일자 2019년10월17일
(30) 우선권주장
62/655,990 2018년04월11일 미국(US)

(71) 출원인
돌비 인터네셔널 에이비
아일랜드 디02 브이케이60 더블린 그랜드 커널 독 랜즈 블록 씨 서 존 로저슨스 키 77
(72) 발명자
데렌티브, 레온
독일 뉘른베르크 90429 도이치헤른스트라쎄 15-19 돌비 저머니 게엠베하 내
페르쉬, 크리스토프
독일 뉘른베르크 90429 도이치헤른스트라쎄 15-19 돌비 저머니 게엠베하 내
피셔, 다니엘
독일 뉘른베르크 90429 도이치헤른스트라쎄 15-19 돌비 저머니 게엠베하 내
(74) 대리인
특허법인 광장리앤고

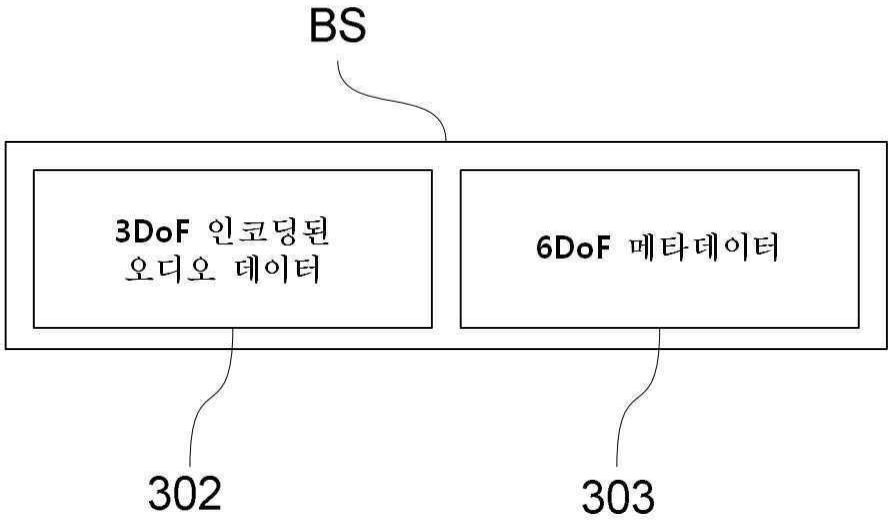
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **6DoF 오디오 렌더링을 위한 방법, 장치 및 시스템, 및 6DoF 오디오 렌더링을 위한 데이터 표현 및 비트스트림 구조**

(57) 요 약

본 개시는, 특히 인코더에서, 오디오 신호를 비트스트림으로 인코딩하는 방법, 장치 및 시스템에 관한 것으로서, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 비트스트림의 하나 이상의 제1 비트스트림 부분으로 인코딩하거나 포함하는 것, 및 6DoF 오디오 렌더링과 연관된 메타데이터를 비트스트림의 하나 이상의 제2 비트스트림 부분으로 인코딩하거나 포함하는 것을 포함한다. 본 개시는 또한, 비트스트림에 기반하여 오디오 신호 디코딩 및 오디오 렌더링하기 위한 방법, 장치 및 시스템에 관한 것이다.

대 표 도 - 도6a



(52) CPC특허분류
G10L 19/24 (2013.01)
H04S 7/303 (2013.01)
H04S 2400/11 (2013.01)

명 세 서

청구범위

청구항 1

비트스트림을 디코딩하기 위한 방법으로서, 상기 방법은:

3 자유도(three degrees of freedom, 3DoF) 오디오 렌더링과 연관된 인코딩된 오디오 신호 데이터 및 6 자유도(six degrees of freedom, 6DoF) 오디오 렌더링과 연관된 메타데이터를 포함하는 상기 비트스트림을 수신하는 단계,

3DoF와 연관된 상기 인코딩된 오디오 신호 데이터를 디코딩하여 3DoF 오디오 신호를 획득하는 단계; 및

상기 3DoF 오디오 신호를 렌더링하여 3DoF 오디오 렌더링 또는 6DoF 오디오 렌더링 중 하나에 기반하는 음장(sound field)을 생성하는 단계 - 상기 6DoF 오디오 렌더링은 상기 3DoF 오디오 신호 및 6DoF와 연관된 상기 메타데이터에 기반하여 6DoF 오디오 신호 데이터를 생성함 - 를 포함하고,

상기 3DoF 오디오 렌더링은:

6DoF와 연관된 상기 메타데이터를 무시(ignore)하는 것; 및

상기 3DoF 오디오 신호를 사용하여 상기 3DoF 오디오 렌더링을 수행하는 것

을 포함하고,

상기 6DoF 오디오 렌더링은:

역변환 함수 A^{-1}을 사용하여 상기 3DoF 오디오 신호로부터 하나 이상의 오디오 소스의 원래의 오디오 신호들 $x*$를 복원하는 것; 및

상기 복원된 원래의 오디오 신호들 $x*$ 및 6DoF와 연관된 상기 메타데이터를 사용하여 상기 6DoF 오디오 렌더링을 수행하는 것

을 포함하는, 방법.

청구항 2

제1항에 있어서, 3DoF와 연관된 상기 인코딩된 오디오 신호 데이터는 변환 함수 A를 사용하여, 상기 하나 이상의 오디오 소스의 원래의 오디오 신호들 x를 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체(spheres) 상에 위치된 대응하는 오디오 객체에 매핑함으로써 생성되는, 방법.

청구항 3

제1항에 있어서, 상기 3DoF 오디오 렌더링과 연관된 상기 인코딩된 오디오 신호 데이터는: 하나 이상의 오디오 객체, 상기 하나 이상의 오디오 객체의 방향성 데이터(directional data) 및 상기 하나 이상의 오디오 객체의 거리 데이터(distance data) 중 적어도 하나를 포함하는, 방법.

청구항 4

제3항에 있어서, 상기 하나 이상의 오디오 객체는 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체 상에 위치되는, 방법.

청구항 5

제1항에 있어서, 6DoF 오디오 렌더링과 연관된 상기 메타데이터는 하나 이상의 디폴트 3DoF 청취자 위치를 나타내는, 방법.

청구항 6

제1항에 있어서, 6DoF 오디오 렌더링과 연관된 상기 메타데이터는: 6DoF 공간의 설명, 하나 이상의 오디오 객체의 오디오 객체 방향들, 가상 현실 환경, 적어도 거리 감쇠, 폐색 및 잔향들 중 적어도 하나에 관련된 매개변수 중 적어도 하나를 나타내는, 방법.

청구항 7

제2항에 있어서, 3DoF 오디오 렌더링과 연관된 상기 인코딩된 오디오 신호 데이터는 하나 이상의 오디오 소스로부터의 상기 원래의 오디오 신호들 및 상기 변환 함수 A에 기반하여 결정되었던, 방법.

청구항 8

제7항에 있어서,

3DoF 오디오 렌더링과 연관된 상기 인코딩된 오디오 신호 데이터는 상기 변환 함수 A를 사용하여 상기 하나 이상의 오디오 소스로부터의 상기 원래의 오디오 신호들을 상기 3DoF 오디오 신호로 변환함으로써 결정되었으며, 상기 변환 함수 A는 상기 하나 이상의 오디오 소스의 상기 원래의 오디오 신호들을 상기 디폴트 3DoF 청취자 위치를 둘러싸는 상기 하나 이상의 구체 상에 위치된 각각의 오디오 객체에 매핑한, 방법.

청구항 9

제1항에 있어서, 상기 비트스트림은 MPEG-H 3D 오디오 표준과 호환되는, 방법.

청구항 10

제1항에 있어서,

상기 3DoF 오디오 렌더링과 연관된 상기 인코딩된 오디오 신호 데이터는 상기 비트스트림의 페이로드의 일부이고,

상기 6DoF 오디오 렌더링과 연관된 상기 메타데이터는 상기 비트스트림의 하나 이상의 확장 컨테이너의 일부인, 방법.

청구항 11

명령어를 포함하는 비일시적 컴퓨터 프로그램 제품으로서, 상기 명령어는 프로세서에 의해 실행될 때, 상기 프로세서로 하여금 제1항의 방법을 실행하게 하는, 비일시적 컴퓨터 프로그램 제품.

청구항 12

비트스트림을 디코딩하기 위한 오디오 디코더 장치로서, 상기 장치는:

3 자유도(3DoF) 오디오 렌더링과 연관된 인코딩된 오디오 신호 데이터 및 6 자유도(6DoF) 오디오 렌더링과 연관된 메타데이터를 포함하는 상기 비트스트림을 수신하기 위한 수신기;

3DoF와 연관된 상기 인코딩된 오디오 신호 데이터를 디코딩하여 3DoF 오디오 신호를 획득하기 위한 디코더; 및

상기 3DoF 오디오 신호를 렌더링하여 3DoF 오디오 렌더링 또는 6DoF 오디오 렌더링 중 하나에 기반하는 음장을 생성하기 위한 렌더러 - 상기 6DoF 오디오 렌더링은 상기 3DoF 오디오 신호 및 6DoF와 연관된 상기 메타데이터에 기반하여 6DoF 오디오 신호 데이터를 생성함 - 를 포함하고,

상기 3DoF 오디오 렌더링은:

6DoF와 연관된 상기 메타데이터를 무시하는 것; 및

상기 3DoF 오디오 신호를 사용하여 상기 3DoF 오디오 렌더링을 수행하는 것

을 포함하고,

상기 6DoF 오디오 렌더링은:

역변환 함수 A^{-1}을 사용하여 상기 3DoF 오디오 신호로부터 하나 이상의 오디오 소스의 원래의 오디오 신호들 $x*$를 복원하는 것; 및

상기 복원된 원래의 오디오 신호들 *x** 및 6DoF와 연관된 상기 메타데이터를 사용하여 상기 6DoF 오디오 렌더링을 수행하는 것

을 포함하는, 오디오 디코더 장치.

발명의 설명

기 술 분 야

[0001] **관련 출원들**

[0002] 본 출원은, 2018년 4월 11일자로 출원된 미국 특허 제62/655,990호의 이익을 주장하고, 이 출원은 그 전문이 본 명세서에 참조로 포함된다.

[0003] **기술 분야**

[0004] 본 개시는 특히, 6 자유도(Six Degrees of Freedom; 6DoF) 오디오 렌더링을 위한 데이터 표현 및 비트스트림 구조와 관련하여 6DoF 오디오 렌더링을 위한 장치, 시스템 및 방법을 제공하는 것에 관한 것이다.

배 경 기 술

[0005] 현재, 사용자의 6DoF 움직임과 결합하여 오디오를 렌더링하기 위한 적절한 해결법이 부족하다. 3 자유도(Three Degrees of Freedom; 3DoF) 움직임(요(yaw), 피치(pitch), 롤(roll))과 결합하여 채널, 객체 및 일차/고차 앰비소닉(First/Higher Order Ambisonics; HOA) 신호들을 렌더링하는 해결법들이 있지만, 사용자의 6DoF(Six Degrees of Freedom) 움직임(요, 피치, 롤 및 병진 운동(translational movement))과 결합된 신호들을 처리하는 것에 대한 지원이 부족하다.

[0006] 일반적으로, 3DoF 오디오 렌더링은, 3DoF 위치로 지칭되는, 사전 결정된 청취자 위치를 둘러싸는 각도 위치들에서 하나 이상의 오디오 소스가 렌더링되는 음장(sound field)을 제공한다. 3DoF 오디오 렌더링의 일 예시가 MPEG-H 3D 오디오 표준(MPEG-H 3DA로 약칭됨)에 포함되어 있다.

[0007] MPEG-H 3DA는 3DoF를 위한 채널, 객체 및 HOA 신호들을 지원하기 위해 개발되었지만, 아직 진정한 6DoF 오디오를 처리할 수 없다. 구상중인(envisioned) MPEG-I 3D 오디오 구현은, 바람직하게는 3DoF 렌더링 하위 호환성을 제공하면서, (바람직하게는 효율적인 신호 생성, 인코딩, 디코딩 및/또는 렌더링을 포함하여) 효율적인 방식으로 3DoF(및 3DoF+) 기능성을 6DoF 3D 오디오 기기로 확장하는 것이 바람직하다.

[0008] 위의 관점에서, 본 개시의 목적은, 예를 들어, MPEG-H 3DA 표준에 따라, 바람직하게는 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 허용하는 3D 오디오 인코딩 및/또는 3D 오디오 렌더링을 위한 방법, 장치 및 데이터 표현 및/또는 비트스트림 구조를 제공하는 것이다.

[0009] 본 개시의 또다른 목적은, 예를 들어, MPEG-H 3DA 표준에 따라, 바람직하게는 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 허용하는 3D 오디오 인코딩 및/또는 3D 오디오 렌더링을 위한 데이터 표현 및/또는 비트스트림 구조와, 예를 들어, MPEG-H 3DA 표준에 따라, 바람직하게는 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 위한 인코딩 및/또는 렌더링 장치를 제공하는 것일 수 있다.

발명의 내용

[0010] 예시적인 양상들에 따르면, 특히 인코더에서, 오디오 신호를 비트스트림으로 인코딩하는 방법이 제공될 수 있고, 본 방법은, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 비트스트림의 하나 이상의 제1 비트스트림 부분으로 인코딩 및/또는 포함하는 단계; 및/또는 6DoF 오디오 렌더링과 연관된 메타데이터를 비트스트림의 하나 이상의 제2 비트스트림 부분으로 인코딩 및/또는 포함하는 단계를 포함한다.

[0011] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는 하나 이상의 오디오 객체의 오디오 신호 데이터를 포함한다.

[0012] 예시적인 양상들에 따르면, 하나 이상의 오디오 객체는 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체(spheres) 상에 위치된다.

[0013] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는 하나 이상의 오디오 객체의 방

향성 데이터(directional data) 및/또는 하나 이상의 오디오 객체의 거리 데이터(distance data)를 포함한다.

[0014] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링과 연관된 메타데이터는 하나 이상의 디폴트 3DoF 청취자 위치를 나타낸다.

[0015] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링과 연관된 메타데이터는, 선택적으로 객체 좌표들을 포함하는, 6DoF 공간의 설명; 하나 이상의 오디오 객체의 오디오 객체 방향들; 가상 현실(virtual reality; VR) 환경; 및/또는 거리 감쇠, 폐색 및/또는 잔향들에 관련된 매개변수들 중 적어도 하나를 포함하거나 나타낸다.

[0016] 예시적인 양상들에 따르면, 본 방법은 하나 이상의 오디오 소스로부터 오디오 신호들을 수신하는 단계; 및/또는 하나 이상의 오디오 소스로부터의 오디오 신호들 및 변환 함수에 기반하여 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 생성하는 단계를 포함한다.

[0017] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는, 변환 함수(transform function)를 사용하여 하나 이상의 오디오 소스로부터의 오디오 신호들을 3DoF 오디오 신호들로 변환함으로써, 생성된다.

[0018] 예시적인 양상들에 따르면, 변환 함수는, 하나 이상의 오디오 소스의 오디오 신호들을, 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체 상에 위치된 각각의 오디오 객체에 매핑하거나 투사한다.

[0019] 예시적인 양상들에 따르면, 본 방법은, 거리 감쇠, 폐색 및/또는 잔향들에 관련된 매개변수들 및/또는 환경적 특성들에 기반하여 변환 함수의 매개변수화(parametrization)를 결정하는 단계를 더 포함할 수 있다.

[0020] 예시적인 양상들에 따르면, 비트스트림은 MPEG-H 3D 오디오 비트스트림 또는 MPEG-H 3D 오디오 신택스를 사용하는 비트스트림이다.

[0021] 예시적인 양상들에 따르면, 비트스트림의 하나 이상의 제1 비트스트림 부분은 비트스트림의 페이로드를 나타내고, 및/또는 상기 하나 이상의 제2 비트스트림 부분은 비트스트림의 하나 이상의 확장 컨테이너를 나타낸다.

[0022] 또 다른 예시적인 양상에 따르면, 특히 디코더 또는 오디오 렌더러에서, 디코딩 및/또는 오디오 렌더링을 위한 방법이 제공될 수 있고, 본 방법은, 비트스트림의 하나 이상의 제1 비트스트림 부분에 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 포함하고 비트스트림의 하나 이상의 제2 비트스트림 부분에 6DoF 오디오 렌더링과 연관된 메타데이터를 더 포함하는 비트스트림을 수신하는 단계, 및/또는 수신된 비트스트림에 기반하여 3DoF 오디오 렌더링 및 6DoF 오디오 렌더링 중 적어도 하나를 수행하는 단계를 포함한다.

[0023] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링을 수행할 때, 3DoF 오디오 렌더링은, 비트스트림의 하나 이상의 제2 비트스트림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터를 폐기하면서, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터에 기반하여 수행된다.

[0024] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링을 수행할 때, 6DoF 오디오 렌더링은, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터 및 비트스트림의 하나 이상의 제2 비트스트림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터에 기반하여 수행된다.

[0025] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는 하나 이상의 오디오 객체의 오디오 신호 데이터를 포함한다.

[0026] 예시적인 양상들에 따르면, 하나 이상의 오디오 객체는 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체 상에 위치된다.

[0027] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는 하나 이상의 오디오 객체의 방향성 데이터 및/또는 하나 이상의 오디오 객체의 거리 데이터를 포함한다.

[0028] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링과 연관된 메타데이터는 하나 이상의 디폴트 3DoF 청취자 위치를 표시한다.

[0029] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링과 연관된 메타데이터는, 선택적으로 객체 좌표들을 포함하는, 6DoF 공간의 설명; 하나 이상의 오디오 객체의 오디오 객체 방향들; 가상 현실(VR) 환경; 및/또는 거리 감쇠, 폐색 및/또는 잔향들에 관련된 매개변수들 중 적어도 하나를 포함하거나 나타낸다.

[0030] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는, 하나 이상의 오디오 소스로부

터의 오디오 신호들 및 변환 함수에 기반하여 생성된다.

[0031] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터는, 변환 함수를 사용하여 하나 이상의 오디오 소스로부터의 오디오 신호들을 3DoF 오디오 신호들로 변환함으로써, 생성된다.

[0032] 예시적인 양상들에 따르면, 변환 함수는, 하나 이상의 오디오 소스의 오디오 신호들을, 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체 상에 위치된 각각의 오디오 객체에 매핑하거나 투사한다.

[0033] 예시적인 양상들에 따르면, 비트스트림은 MPEG-H 3D 오디오 비트스트림 또는 MPEG-H 3D 오디오 신택스를 사용하는 비트스트림이다.

[0034] 예시적인 양상들에 따르면, 비트스트림의 하나 이상의 제1 비트스트림 부분은 비트스트림의 페이로드를 나타내고, 및/또는 하나 이상의 제2 비트스트림 부분은 비트스트림의 하나 이상의 확장 컨테이너를 나타낸다.

[0035] 예시적인 양상들에 따르면, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터 및 비트스트림의 하나 이상의 제2 비트스트림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터에 기반하여, 6DoF 오디오 렌더링을 수행하는 것은, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터 및 역변환 함수에 기반하여 6DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 생성하는 단계를 포함한다.

[0036] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링과 연관된 오디오 신호 데이터는, 6DoF 오디오 렌더링과 연관된 메타데이터 및 역변환 함수를 사용하여 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 변환함으로써, 생성된다.

[0037] 예시적인 양상들에 따르면, 역변환 함수는, 하나 이상의 오디오 소스의 오디오 신호들을, 디폴트 3DoF 청취자 위치를 둘러싸는 하나 이상의 구체 상에 위치된 각각의 오디오 객체에 매핑하거나 투사하는 변환 함수의 역함수이다.

[0038] 예시적인 양상들에 따르면, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터에 기반하여, 3DoF 오디오 렌더링을 수행하는 것은, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터 및 비트스트림의 하나 이상의 제2 비트스트림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터에 기반하여, 디폴트 3DoF 청취자 위치에서, 6DoF 오디오 렌더링을 수행하는 것과 동일한 음장 생성을 초래한다.

[0039] 또 다른 예시적인 양상에 따르면, 오디오 렌더링을 위한 비트스트림이 제공될 수 있고, 이 비트스트림은 비트스트림의 하나 이상의 제1 비트스트림 부분에 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 포함하고 비트스트림의 제2 비트스트림 부분에 6DoF 오디오 렌더링과 연관된 메타데이터를 더 포함한다. 이 양상은 위의 예시적인 양상 중 임의의 하나 이상과 결합될 수 있다.

[0040] 또 다른 예시적인 양상에 따르면, 프로세서를 포함하는 장치, 특히 인코더가 제공될 수 있고, 이 프로세서는 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 비트스트림의 하나 이상의 제1 비트스트림 부분으로 인코딩 및/또는 포함하고; 6DoF 오디오 렌더링과 연관된 메타데이터를 비트스트림의 하나 이상의 제2 비트스트림 부분으로 인코딩 및/또는 포함하고; 및/또는 인코딩된 비트스트림을 출력하도록 구성된다. 이 양상은 위의 예시적인 양상 중 임의의 하나 이상과 결합될 수 있다.

[0041] 또 다른 예시적인 양상에 따르면, 프로세서를 포함하는 장치, 특히 디코더 또는 오디오 렌더러가 제공될 수 있고, 이 프로세서는 비트스트림의 하나 이상의 제1 비트스트림 부분에 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 포함하고 비트스트림의 하나 이상의 제2 비트스트림 부분에 6DoF 오디오 렌더링과 연관된 메타데이터를 더 포함하는 비트스트림을 수신하고, 및/또는 수신된 비트스트림에 기반하여 3DoF 오디오 렌더링 및 6DoF 오디오 렌더링 중 적어도 하나를 수행하도록 구성된다. 이 양상은 위의 예시적인 양상 중 임의의 하나 이상과 결합될 수 있다.

[0042] 예시적인 양상들에 따르면, 3DoF 오디오 렌더링을 수행할 때, 프로세서는, 비트스트림의 하나 이상의 제2 비트스트림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터를 폐기하면서, 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터에 기반하여 3DoF 오디오 렌더링을 수행하도록 구성된다.

[0043] 예시적인 양상들에 따르면, 6DoF 오디오 렌더링을 수행할 때, 프로세서는 비트스트림의 하나 이상의 제1 비트스트림 부분에서의 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터 및 비트스트림의 하나 이상의 제2 비트스트

림 부분에서의 6DoF 오디오 렌더링과 연관된 메타데이터에 기반하여 6DoF 오디오 렌더링을 수행하도록 구성된다.

[0044] 또 다른 예시적인 양상에 따르면, 프로세서에 의해 실행될 때, 프로세서로 하여금, 특히 인코더에서, 오디오 신호를 비트스트림으로 인코딩하는 방법을 실행하게 하는 명령어들을 포함하는 비-일시적 컴퓨터 프로그램 제품이 제공될 수 있고, 이 방법은, 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 비트스트림의 하나 이상의 제1 비트스트림 부분으로 인코딩 또는 포함하는 단계; 및/또는 6DoF 오디오 렌더링과 연관된 메타데이터를 비트스트림의 하나 이상의 제2 비트스트림 부분으로 인코딩 또는 포함하는 단계를 포함한다. 이 양상은 위의 예시적인 양상 중 임의의 하나 이상과 결합될 수 있다.

[0045] 또 다른 예시적인 양상에 따르면, 프로세서에 의해 실행될 때, 프로세서로 하여금, 특히 디코더 또는 오디오 렌더러에서, 디코딩 및/또는 오디오 렌더링을 위한 방법을 실행하게 하는 명령어들을 포함하는 비-일시적 컴퓨터 프로그램 제품이 제공될 수 있고, 이 방법은, 비트스트림의 하나 이상의 제1 비트스트림 부분에 3DoF 오디오 렌더링과 연관된 오디오 신호 데이터를 포함하고 비트스트림의 하나 이상의 제2 비트스트림 부분에 6DoF 오디오 렌더링과 연관된 메타데이터를 더 포함하는 비트스트림을 수신하는 단계, 및/또는 수신된 비트스트림에 기반하여 3DoF 오디오 렌더링 및 6DoF 오디오 렌더링 중 적어도 하나를 수행하는 단계를 포함한다. 이 양상은 위의 예시적인 양상 중 임의의 하나 이상과 결합될 수 있다.

[0046] 본 개시의 추가적 양상들은 대응하는 컴퓨터 프로그램들 및 컴퓨터 판독 가능한 저장 매체에 관한 것이다.

[0047] 방법의 단계들 및 장치의 특징들은 다수의 방식으로 상호 교환적일 수 있음이 이해될 것이다. 특히, 개시된 방법의 세부 사항은 통상의 기술자가 이해하는 바와 같이, 방법의 일부 또는 전부 또는 단계들을 실행하도록 적응된 장치로서 구현될 수 있고, 그 반대도 마찬가지이다. 특히, 방법에 관한 각각의 진술은 대응하는 장치에 마찬가지로 적용되며, 그 반대도 마찬가지이다.

도면의 간단한 설명

[0048] 본 개시의 예시적인 실시예들이 첨부 도면들을 참조하여 아래에 설명되고, 여기서 동일한 참조 부호는 동일하거나 유사한 요소를 나타낼 수 있고, 여기서:

도 1은 본 개시의 예시적인 양상들에 따른 MPEG-H 3D 오디오 디코더/인코더 인터페이스를 포함하는 예시적인 시스템을 개략적으로 도시한다.

도 2는 방(room)(6DoF 공간)의 6DoF 장면의 예시적인 평면도(top view)를 개략적으로 도시한다.

도 3은 도 2의 6DoF 장면 및 본 개시의 예시적인 양상들에 따른 3DoF 오디오 데이터 및 6DoF 확장 메타데이터의 예시적인 평면도를 개략적으로 도시한다.

도 4a는 본 개시의 예시적인 양상들에 따라 3DoF, 6DoF 및 오디오 데이터를 처리하기 위한 예시적인 시스템을 개략적으로 도시한다.

도 4b는 본 개시의 예시적인 양상들에 따른 6DoF 오디오 렌더링 및 3DoF 오디오 렌더링을 위한 예시적인 디코딩 및 렌더링 방법을 개략적으로 도시한다.

도 5는 도 2 내지 도 4b 중 하나 이상에 따른 시스템의 3DoF 위치에서 6DoF 오디오 렌더링 및 3DoF 오디오 렌더링의 예시적인 매칭 조건을 개략적으로 도시한다.

도 6a는 본 개시의 예시적인 양상들에 따른 예시적인 데이터 표현 및/또는 비트스트림 구조를 개략적으로 도시한다.

도 6b는 본 개시의 예시적인 양상들에 따른 도 6a의 데이터 표현 및/또는 비트스트림 구조에 기반하는 예시적인 3DoF 오디오 렌더링을 개략적으로 도시한다.

도 6c는 본 개시의 예시적인 양상들에 따른 도 6a의 데이터 표현 및/또는 비트스트림 구조에 기반하는 예시적인 6DoF 오디오 렌더링을 개략적으로 도시한다.

도 7a는 본 개시의 예시적인 양상들에 따른 3DoF 오디오 신호 데이터에 기반하는 6DoF 오디오 인코딩 변환 A를 개략적으로 도시한다.

도 7b는 본 개시의 예시적인 양상들에 따라 3DoF 오디오 신호 데이터에 기반하여 6DoF 오디오 신호 데이터를 근

사/복원(approximating/restoring)하기 위한 6DoF 오디오 디코딩 변환 A^{-1}을 개략적으로 도시한다.

도 7c는, 본 개시의 예시적인 양상들에 따른, 도 7b의 근사/복원된 6DoF 오디오 신호 데이터에 기반하는 예시적인 6DoF 오디오 렌더링을 개략적으로 도시한다.

도 8은 본 개시의 예시적인 양상들에 따른 3DoF/6DoF 비트스트림 인코딩 방법의 예시적인 흐름도를 개략적으로 도시한다.

도 9는 본 개시의 예시적인 양상들에 따른 3DoF 및/또는 6DoF 오디오 렌더링의 방법들의 예시적인 흐름도를 개략적으로 도시한다.

발명을 실시하기 위한 구체적인 내용

[0049] 다음에서, 바람직한 예시적인 양상들이 첨부 도면들을 참조하여 더 상세히 설명될 것이다. 상이한 도면들 및 실시예들에서 동일하거나 유사한 특징은 유사한 참조 부호로 지칭될 수 있다. 다양한 바람직한 예시적인 양상들에 관련된 아래의 상세한 설명이 본 발명의 범위를 제한하는 것으로 의미되지 않는다는 점을 이해해야 한다.

[0050] 본 명세서에 사용된 바와 같이, "MPEG-H 3D 오디오"는, ISO/IEC 23008-3에서 및/또는 ISO/IEC 23008-3 표준의 임의의 과거 및/또는 향후 보정본, 편집본 또는 기타 버전에서 표준화된 사양을 참조할 것이다.

[0051] 본 명세서에 사용된 바와 같이, MPEG-I 3D 오디오 구현은, 바람직하게는 3DoF 렌더링 하위 호환성을 제공하면서, 3DoF(및 3DoF+) 기능성을 6DoF 3D 오디오로 확장하는 것이 바람직하다.

[0052] 본 명세서에 사용된 바와 같이, 3DoF는 전형적으로 3개의 매개변수(예를 들어, 요, 피치, 롤)로 지정된 사용자의 머리 움직임, 특히 머리 회전을 정확하게 처리할 수 있는 시스템이다. 이러한 시스템들은 종종 가상 현실(VR)/증강 현실(augmented reality; AR)/혼합 현실(mixed reality; MR) 시스템 또는 기타 이러한 유형의 음향 환경과 같은 다양한 게임 시스템(gaming system)들에 이용 가능하다.

[0053] 본 명세서에 사용된 바와 같이, 6DoF는 전형적으로 3DoF 및 병진 운동을 정확하게 처리할 수 있는 시스템이다.

[0054] 본 개시의 예시적인 양상들은, 오디오 시스템(예를 들어, MPEG-I 오디오 표준과 호환되는 오디오 시스템)에 관한 것이고, 여기서 오디오 렌더러는, 관련 메타데이터를 MPEG 표준(예를 들어, MPEG-H 3DA 표준)과 호환되는 오디오 렌더러 입력 포맷과 같은 3DoF 포맷으로 변환함으로써, 6DoF로 기능성을 확장한다.

[0055] 도 1은, 6DoF 경험을 사용가능하게 하기 위해, 기존의 3DoF 시스템에 추가하여 메타데이터 확장 및/또는 오디오 렌더러 확장을 사용하도록 구성되는 예시적인 시스템(100)을 도시한다. 시스템(100)은 (예시적으로 하나 이상의 오디오 소스(101a)를 포함할 수 있는) 원래의 환경(101), 콘텐츠 포맷(102)(예를 들어, 3D 오디오 데이터를 포함하는 비트스트림), 인코더(103) 및 제안 메타데이터 인코더 확장(106)을 포함한다. 시스템(100)은 또한, 3D 오디오 렌더러(105)(예를 들어, 3DoF 렌더러) 및 제안 렌더러 확장(107)(예를 들어, 재현된 환경(108)에 대한 6DoF 렌더러 확장)을 포함할 수 있다.

[0056] 3DoF를 갖는 3D 오디오 렌더링 방법에서, 사전 결정된 3DoF 위치에서 사용자의 각도 방향의 각도(예를 들어, 요 각도 y, 피치 각도 p, 롤 각도 r)만이 3DoF 오디오 렌더러(105)에 입력될 수 있다. 확장된 6DoF 기능성을 갖는 사용자의 위치 좌표(예를 들어, x, y 및 z)가 추가적으로 6DoF 오디오 렌더러(확장 렌더러)에 입력될 수 있다.

[0057] 본 개시의 이점은 인코더와 디코더 사이에 전송되는 비트스트림의 비트 레이트 향상을 포함한다. 비트 스트림은 MPEG-I 오디오 표준 및/또는 MPEG-H 3D 오디오 표준과 같은 표준을 준수하여 인코딩 및/또는 디코딩될 수 있고, 적어도 MPEG-H 3D 오디오 표준과 같은 표준과 하위 호환될 수 있다.

[0058] 일부 예시에서, 본 개시의 예시적인 양상들은, 복수의 시스템과 호환되는 단일 비트스트림(예를 들어, MPEG-H 3D 오디오(3DA) 비트스트림(BS), 또는 MPEG-H 3DA BS의 신택스를 사용하는 비트스트림)의 처리에 관한 것이다.

[0059] 예를 들어, 일부 예시적인 양상에서, 오디오 비트스트림은, 2개 이상의 상이한 렌더러, 예를 들어, 하나의 표준(예를 들어, MPEG-H 3D 오디오 표준)과 호환될 수 있는 3DoF 오디오 렌더러, 및 다른 제2 표준(예를 들어, MPEG-I 오디오 표준)과 호환될 수 있는 신규 정의된 6DoF 오디오 렌더러 또는 렌더러 확장과 호환될 수 있다.

[0060] 본 개시의 예시적인 양상들은, 바람직하게는 동일한 오디오 출력을 생성하기 위해, 동일한 오디오 비트스트림에 대한 디코딩 및 렌더링을 수행하도록 구성되는 상이한 디코더들에 관한 것이다.

[0061] 예를 들어, 본 개시의 예시적인 양상들은 동일한 비트스트림(예를 들어, 3DA BS, 또는 3DA BS를 사용하는 비트

스트림)에 대해 동일한 출력을 생성하도록 구성되는 3DoF 디코더 및/또는 3DoF 렌더러 및/또는 6DoF 디코더 및/또는 6DoF 렌더러에 관한 것이다. 예시적으로, 비트스트림은, VR/AR/MR(가상 현실/증강 현실/혼합 현실) 공간에서 청취자의 정의된 위치들에 관한 정보를, 예를 들어, 6DoF 메타데이터의 일부로서, 포함할 수 있다.

[0062] 본 개시는 또한, 예시적으로 (예를 들어, MPEG-I 오디오 환경과 호환되는) 6DoF 정보를 각각 인코딩 및/또는 디코딩하도록 구성되는 인코더들 및/또는 디코더들에 관한 것이고, 본 개시의 이러한 인코더들 및/또는 디코더들은, 다음의 이점:

[0063] • VR/AR/MR 관련 오디오 데이터의 품질-효율적인 및 비트 레이트-효율적인 표현, 및 그의 오디오 비트스트림 신택스로의 캡슐화(예를 들어, MPEG-H 3D 오디오 BS);

[0064] • 다양한 시스템들 사이의 하위 호환성(예를 들어, MPEG-H 3DA 표준 및 구상중인 MPEG-I 오디오 표준) 중 하나 이상을 제공한다.

[0065] 바람직하게는 3DoF-해결법과 6DoF-해결법 사이의 경쟁을 회피하고 현재와 향후 기술 사이의 원활한 이행(smooth transition)을 제공하기 위해, 하위 호환성은 매우 유익하다.

[0066] MPEG-I 오디오와 같은 6DoF 오디오 시스템에, MPEG-H 3D 오디오와 같은 3DoF 오디오 시스템으로의 하위 호환성을 제공하는 것과 같은, 예를 들어, 3DoF 오디오 시스템과 6DoF 오디오 시스템 사이의 하위 호환성은, 매우 유익할 수 있다.

[0067] 본 개시의 예시적인 양상들에 따르면, 이는:

[0068] • 3DoF *오디오 자료* 코딩된 데이터 및 관련 메타데이터; 및

[0069] • 6DoF 관련 *메타데이터*

[0070] 로 구성되는 6DoF 관련 시스템에 대해, 예를 들어, 비트스트림 레벨에서, 하위 호환성을 제공함으로써 실현될 수 있다.

[0071] 본 개시의 예시적인 양상들은, 제1 유형의 오디오 비트스트림(예를 들어, MPEG-H 3DA BS) 신택스와 같은 표준 3DoF 비트스트림 신택스 - 이는 제1 유형의 오디오 비트스트림(예를 들어, MPEG-H 3DA BS)의 하나 이상의 확장 컨테이너에 MPEG-I 오디오 비트스트림 요소들과 같은 6DoF 비트스트림 요소들을 캡슐화함 - 에 관한 것이다.

[0072] 실행 레벨에서 하위 호환성을 보장하는 시스템을 제공하기 위해, 다음의 시스템들 및/또는 구조들이 관련될 수 있고, 발생될 수 있다:

[0073] 1a. 3DoF 시스템(예를 들어, MPEG-H 3DA의 표준들과 호환되는 시스템들)은, 모든 6DoF 관련 신택스 요소들을 *무시(ignore)*(예를 들어, MPEG-H 3D 오디오 비트스트림 신택스의 "mpegh3daExtElement()" 또는 "mpegh3daExtElementConfig()"의 기능성에 기반하여 MPEG-I 오디오 비트스트림 신택스 요소들을 무시)*할 수 있을 것이고,* 즉 3DoF 시스템(디코더/렌더러)은 바람직하게는 (예를 들어, 6DoF 관련 데이터 및/또는 메타데이터를 판독하지 않음으로써) 추가적인 6DoF 관련 데이터 및/또는 메타데이터를 방치(neglect)하도록 구성될 수 있고;

[0074] 2a. 비트스트림 페이로드의 나머지 부분(예를 들어, MPEG-H 3DA 비트스트림 파서(parser)와 호환되는 데이터 및/또는 메타데이터를 포함하는 MPEG-I 오디오 비트스트림 페이로드)은, 원하는 오디오 출력을 생성하기 위해, *3DoF* 시스템(예를 들어, 레거시 MPEG-H 3DA 시스템)*에 의해 디코딩*될 수 있을 것이고, 즉 3DoF 시스템(디코더/렌더러)은 바람직하게는 BS의 3DoF 부분을 디코딩하도록 구성될 수 있고;

[0075] 3a. 6DoF 시스템(예를 들어, MPEG-I 오디오 시스템)은, 오디오 비트스트림의 3DoF 관련 및 6DoF 관련 부분 모두를 처리하고, VR/AR/MR 공간 내의 하위 호환되는 사전 정의된 3DoF 위치(들)에서, 3DoF 시스템(예를 들어, MPEG-H 3DA 시스템들)의 오디오 출력과 *매치*되는 오디오 출력을 생성할 수 있을 것이고, 즉 6DoF 시스템(디코더/렌더러)은 바람직하게는 디폴트 3DoF 위치(들)에서, 3DoF 렌더링되는 음장/오디오 출력과 매치되는 음장/오디오 출력을 렌더링하도록 구성될 수 있고;

[0076] 4a. 6DoF 시스템(예를 들어, MPEG-I 오디오 시스템)은, 하위 호환되는 사전 정의된 3DoF 위치(들) 주위에서 오디오 출력의 *원활한* 변화(천이)를 제공(즉, 6DoF 공간에서 연속되는 음장을 제공)할 것이고, 즉 6DoF 시스템(디코더/렌더러)은 바람직하게는 디폴트 3DoF 위치(들)에서 3DoF 렌더링되는 음장/오디오 출력으로 원활하게 천이

되는 음장/오디오 출력을, 디폴트 3DoF 위치(들)의 주변에서, 렌더링하도록 구성될 수 있다.

[0077] 일부 예시에서, 본 개시는, 하나, 그 이상의 또는 일부 3DoF 위치(들)에서, 3DoF 오디오 렌더러(예를 들어, MPEG-H 3D 오디오 렌더러)와 동일한 오디오 출력을 생성하는 6DoF 오디오 렌더러(예를 들어, MPEG-I 오디오 렌더러)를 제공하는 것에 관한 것이다.

[0078] 현재, 6DoF 오디오 시스템에 바로 3DoF 관련 오디오 신호들 및 메타데이터를 직접 전송할 때, 단점들이 존재하며, 이는 다음의 것:

[0079] 1. 비트 레이트 증가(즉, 6DoF 관련 오디오 신호들 및 메타데이터에 추가되어 3DoF 관련 오디오 신호들 및 메타데이터가 전송됨); 및

[0080] 2. 제한된 유효성(즉, 3DoF 관련 오디오 신호(들) 및 메타데이터는 3DoF 위치(들)에 대해서만 유효함)을 포함한다.

[0081] 본 개시의 예시적인 양상들은 위의 단점들을 극복하는 것에 관한 것이다.

[0082] 일부 예시에서, 본 개시는 다음의 것:

[0083] 1. 원래의 오디오 소스 신호들 및 메타데이터 대신에(또는 이에 무료의 추가로서) 3DoF 호환되는 오디오 신호(들) 및 메타데이터(예를 들어, MPEG-H 3D 오디오와 호환되는 신호들 및 메타데이터)를 사용하는 것; 및/또는

[0084] 2. 고수준의 음장 근사치를 유지하면서, 3DoF 위치(들)에서 (콘텐츠 창작자에 의해 정의된) 6DoF 공간으로의 적용 가능성(6DoF 렌더링의 사용)의 범위를 증가시키는 것에 관한 것이다.

[0085] 본 개시의 예시적인 양상들은 이러한 목표를 충족시키고 6DoF 렌더링 기능성을 제공하기 위해 이러한 신호(들)를 효율적으로 생성, 인코딩, 디코딩 및 렌더링하는 것에 관한 것이다.

[0086] 도 2는 예시적인 방(201)의 예시적인 평면도(202)를 도시한다. 도 2에 도시된 바와 같이, 예시적인 청취자는 수개의 오디오 소스 및 특별한 벽 기하학적 구조(non-trivial wall geometries)를 가진 방 가운데에 서 있다. 6DoF 기기들(예를 들어, 6DoF 능력을 제공하는 시스템들)에서, 예시적인 청취자는 돌아다닐 수 있지만, 일부 예시에서 디폴트 3DoF 위치(206)는 (예를 들어, 콘텐츠 창작자의 의도 또는 설정에 따른) 최상의 VR/AR/MR 오디오 경험의 의도된 영역에 대응할 수 있다고 가정된다.

[0087] 특히, 예시적으로 도 2는 벽들(203), 6DoF 공간(204), 예시적인(선택적) 방향성 벡터(205)(예를 들어, 하나 이상의 음원(sound source)이 방향성 있게 사운드를 방출하는 경우), 3DoF 청취자 위치(206)(디폴트 3DoF 위치(206)) 및 도 2에 별 형상으로 예시적으로 도시된 오디오 소스들(207)을 도시한다.

[0088] 도 3은 예를 들어 도 2에서와 같은 예시적인 6DoF VR/AR/MR 장면뿐만 아니라 (예를 들어, MPEG-H 3D 오디오 비트스트림과 같은) 3DoF 오디오 비트스트림(302)에 포함된 오디오 객체들(오디오 데이터 + 메타데이터)(320) 및 확장 컨테이너(303)를 도시한다. 오디오 비트스트림(302) 및 확장 컨테이너(303)는 MPEG 표준(예를 들어, MPEG-H 또는 MPEG-I)과 호환되는 장치 또는 시스템을 통해(예를 들어, 소프트웨어, 하드웨어 또는 클라우드를 통해) 인코딩될 수 있다.

[0089] 본 개시의 예시적인 양상들은, 6DoF 오디오 렌더러(예를 들어, MPEG-I 오디오 렌더러)를 사용할 때, 3DoF 오디오 렌더러(예를 들어, MPEG-H 오디오 렌더러) 출력 신호(이는 물리적 법칙 사운드 전파와 일치하거나 일치하지 않을 수 있음)에 대응하는 방식으로, "3DoF 위치"에서 음장을 재생성하는 것에 관한 것이다. 이 음장은, 바람직하게는 원래의 "오디오 소스들"에 기반해야 하고, 대응하는 VR/AR/MR 환경(예를 들어, "벽들"의 효과, 구조, 사운드 반사, 잔향 및/또는 폐색 등)의 복잡한 기하학적 구조의 영향을 반영해야 한다.

[0090] 본 개시의 예시적인 양상들은, 인코더에 의한, 전술된 하나, 그 이상의, 또는 바람직하게는 모든 대응하는 요건(1a) 내지 (4a)의 충족을 보장하는 방식으로 이 시나리오를 설명하는 모든 관련 정보의 매개변수화에 관한 것이다.

[0091] 2개의 오디오 렌더링 모드(즉, 3DoF 및 6DoF)가 병렬로 실행되고 보간 알고리즘이 6DoF 공간에서 대응하는 출력에 적용되면, 이러한 접근법은 차선적일 것인데, 그 이유는:

[0092] • 2개의 구분된 렌더링 알고리즘(즉, 하나는 특정 3DoF 위치들에 대한 것이고, 하나는 6DoF 공간에 대한 것임)의 병렬 실행;

[0093] • (3DoF 오디오 렌더러용으로 추가적인 오디오 데이터 전송을 위해서) 다량의 오디오 데이터

[0094] 를 필요로 하기 때문이다.

[0095] 본 개시의 예시적인 양상들은, 바람직하게는 (예를 들어, 2개의 오디오 렌더링 모드의 병렬 실행 대신에) 단일 오디오 렌더링 모드만이 실행되고, 및/또는 3DoF 오디오 데이터가 바람직하게는 (예를 들어, 3DoF 오디오 데이터 및 원래의 음원(들) 데이터를 전송하는 대신에) 원래의 음원(들) 신호(들)를 복원 및/또는 근사하기 위한 추가적인 메타데이터를 갖는 6DoF 오디오 렌더링에 사용된다는 점에서, 위의 단점들을 회피한다.

[0096] 본 개시의 예시적인 양상들은, (1) 바람직하게는 특정 위치(들)에서 (예를 들어, MPEG-H 3DA와 호환되는) 3DoF 오디오 렌더링 알고리즘과 정확히 동일한 출력을 생성하는 (예를 들어, MPEG-I 오디오와 호환되는) 단일 6DoF 오디오 렌더링 알고리즘, 및/또는 (2) 6DoF 오디오 비트스트림 데이터(예를 들어, MPEG-I 오디오 비트스트림 데이터)의 3DoF-관련 및 VR/AR/MR-관련 부분에서 용장성(redundancy)을 감소시키기 위해, 오디오(예를 들어, 3DoF 오디오 데이터) 및 6DoF 관련 오디오 메타데이터를 나타내는 것에 관한 것이다.

[0097] 본 개시의 예시적인 양상들은:

[0098] • (예를 들어, 3DoF 오디오 비트스트림 신택스의 코어 부분(core part)에서의), 바람직하게는 3DoF 오디오 시스템에 의해 디코딩되며, 바람직하게는 (디폴트) 3DoF 위치(들)에서 원하는 음장을 충분히 근사하는, 오디오 소스 신호들 및 메타데이터를 전송하고;

[0099] • (예를 들어, 3DoF 오디오 비트스트림 신택스의 확장 부분에서의), 6DoF 오디오 렌더링을 위해 원래의 오디오 소스 신호들을 근사(복원)하기 위해 사용되는, 6DoF 관련 메타데이터 및/또는 추가적 데이터(예를 들어, 파라메트릭 또는/및 신호 데이터)를 전송하도록,

[0100] 제2 표준화된 포맷 비트스트림(예를 들어, MPEG-I와 같은 향후 표준) 또는 그의 부분들 및 6DoF 관련 메타데이터를 캡슐화하기 위해, 제1 표준화된 포맷 비트스트림(예를 들어, MPEG-H 3DA BS) 신택스를 사용하는 것에 관한 것이다.

[0101] 본 개시의 양상은 인코더 측에서 원하는 "3DoF 위치(들)" 및 3DoF 오디오 시스템(예를 들어, MPEG-H 3DA 시스템) 호환되는 신호들의 결정에 관한 것이다.

[0102] 예를 들어, 도 3과 관련하여 도시된 바와 같이, 일부 3DoF 시스템(예컨대, MPEG-H 3DA 시스템)이 VR/AR/MR 환경적 효과들(예를 들어, 폐색, 잔향 등)을 고려할 수 없기 때문에, 3DA를 위한 가상(virtual) 3DA 객체 신호들이 (*신호 x_{3DA}에 기반하여*) 특정 3DoF 위치에서 동일한 음장을 생성할 수 있고, 이는 바람직하게는 특정 3DoF 위치(들)("웨트(wet)" 신호들)에 대한 VR 환경의 효과들을 포함해야 한다. 도 3에 도시된 방법 및 프로세스는 다양한 시스템 및/또는 제품을 통해 수행될 수 있다.

[0103] 역함수 A^{-1}는, 일부 예시적인 양상에서, (VR 환경의 효과들로부터 자유로운) 원래의 "드라이(dry)" 신호들 x를 근사하기 위해 필요하므로, 바람직하게는 (즉, VR 환경의 효과들을 제거하는) "언-웨트(un-wet)"의 이러한 신호들은 양호해야 한다.

[0104] 3DoF 렌더링을 위한 오디오 신호(들)((x_{3DA}))은 바람직하게는 3DoF 및 6DoF 오디오 렌더링 모두에 대해 동일한/유사한 출력을 제공하도록, 다음에 기반하여 정의될 수 있다:

[0105] [수학식 1]

[0106] $$F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}(x) \; for \; 3DoF$$

[0107] 오디오 객체들은 표준화된 비트 스트림에 포함될 수 있다. 이 비트 스트림은 MPEG-H 3DA 및/또는 MPEG-I와 같은 다양한 표준들을 준수하여 인코딩될 수 있다.

[0108] BS는 객체 신호들, 객체 방향들 및 객체 거리들에 관한 정보를 포함할 수 있다.

[0109] 도 3은, 예를 들어, BS에 확장 메타데이터를 포함할 수 있는 확장 컨테이너(303)를 예시적으로 더 도시한다. BS의 확장 컨테이너(303)는, 다음의 메타데이터: (i) 3DoF (디폴트) 위치 매개변수들; (ii) 6DoF 공간 설명 매개

변수들(객체 좌표들); (iii) (선택적) 객체 방향성 매개변수들; (iv) (선택적) VR/AR/MR 환경 매개변수들; 및/또는 (v) (선택적) 거리 감쇠 매개변수들, 폐색 매개변수들 및/또는 잔향 매개변수들 등 중 적어도 하나를 포함할 수 있다.

[0110] 다음에 기반하여 원하는 오디오 렌더링의 근사치가 포함될 수 있다:

[0111] [수학식 2]

[0112] $$F_{6DoF}(x^*) \approx F_{6DoF}(x)\ for\ 6DoF$$

[0113] 근사치는 VR 환경에 기반할 수 있고, 환경 특성들은 확장 컨테이너 메타데이터에 포함될 수 있다.

[0114] 추가적으로 또는 선택적으로, 6DoF 오디오 렌더러(예를 들어, MPEG-I 오디오 렌더러) 출력에 대한 평활화가, 바람직하게는 다음에 기반하여 제공될 수 있다:

[0115] [수학식 3]

[0116] $F_{6DoF} \subset G^{i\geq 0}\ for\ 3DoF +$, $G^{i\geq 0}$ - 기하학적 연속 클래스

[0117] 본 개시의 예시적인 양상들은, 바람직하게는 다음에 기반하여, 인코더 측에서, 3DoF 오디오 객체들(예를 들어 MPEG-H 3DA 객체들)을 정의하는 것에 관한 것이다:

[0118] [수학식 4]

[0119] $$x_{3DA} := A(x), \left\| F_{3DoF}(x_{3DA}) - F_{6DoF}\ (x)\ for\ 3DoF \right\| \rightarrow min$$

[0120] 본 개시의 양상은 다음에 기반하여 디코더 상에서 원래의 객체들을 복구하는 것에 관한 것이다:

[0121] [수학식 5]

[0122] $$x^* := A^{-1}(x_{3DA})$$

[0123] 여기서, x는 음원/객체 신호에 관한 것이고, x^*는 음원/객체 신호들의 근사치에 관한 것이고, $F(x)\ for\ 3DoF\ /\ for\ 6DoF$는 3DoF/6DoF 청취자 위치(들)에 대한 오디오 렌더링 함수에 관한 것이고, 3DoF는 주어진 참조 호환성 위치(들) ∈ 6DoF 공간에 관한 것이고; 6DoF는 임의의 허용된 위치(들) ∈ VR 장면에 관한 것이며;

[0124] • $F_{6FoF}(x)$는 디코더 지정된 6DoF 오디오 렌더링(예를 들어, MPEG-I 오디오 렌더링)에 관한 것이고;

[0125] • $F_{3DoF}(x_{3DA})$는 디코더 지정된 3DoF 렌더링(예를 들어, MPEG-H 3DA 렌더링)에 관한 것이고;

[0126] • A, A^{-1}은 신호들 x와 그 역(A^{-1})에 기반하여 신호들 x_{3DA}를 근사하는 함수(A)에 관한 것이다.

[0127] 근사된 음원/객체 신호들은 바람직하게는 3DoF 오디오 렌더러 출력 신호에 대응하는 방식으로 "3DoF 위치"에서 6DoF 오디오 렌더러를 사용하여 재생성된다.

[0128] 음원/객체 신호들은 바람직하게는 원래의 "오디오 소스들"에 기반하고 대응하는 VR/AR/MR 환경(예를 들어, "벽", 구조, 잔향, 폐색 등)의 복잡한 기하학적 구조의 영향을 반영하는 음장에 기반하여 근사된다.

[0129] 즉, 3DA에 대한 가상 3DA 객체 신호들은 바람직하게는 (*신호 x_{3DA}에 기반하여*) 특정 3DoF 위치에서 동일한 음장을 생성하고, 이는 특정 3DoF 위치(들)에 대한 VR 환경의 효과들을 포함한다.

[0130] 다음의 것:

[0131] • 3DoF 오디오 렌더링을 위한 오디오 신호(들): x_{3DA}

[0132] • 3DoF 또는 6DoF 오디오 렌더링 기능성:

[0133] [수학식 6]

[0134] $F_{3DoF}(x_{3DA})$ 또는 $F_{6DoF}(x)$ 중 어느 하나는,

[0135] 렌더링 측에서(예를 들어, MPEG-H 또는 MPEG-I 표준과 같은 표준을 준수하는 디코더에 의해) 이용 가능할 수 있다.

[0136] 6DoF 오디오 렌더링의 경우, 추가적으로 6DoF 오디오 렌더링 기능성을 위한 6DoF 메타데이터가 (예를 들어, 3DoF 오디오 신호 x_{3DA} 및 6DoF 메타데이터에 기반하여, 예를 들어, 하나 이상의 오디오 소스의 오디오 신호들 x를 근사/복원하기 위해) 렌더링 측에서 이용 가능할 수 있다.

[0137] 본 개시의 예시적인 양상들은, (i) 3DoF 오디오 객체들(예를 들어, MPEG-H 3DA 객체들)의 정의 및/또는 (ii) 원래의 오디오 객체들의 복구(근사)에 관한 것이다.

[0138] 오디오 객체들은 예시적으로 3DoF 오디오 비트스트림(예컨대, MPEG-H 3DA BS)에 포함될 수 있다.

[0139] 비트스트림은 객체 오디오 신호들, 객체 방향들 및/또는 객체 거리들에 관한 정보를 포함할 수 있다.

[0140] (예를 들어, MPEG-H 3DA BS와 같은 비트스트림의) 확장 컨테이너는, 다음의 메타데이터: (i) 3DoF(디폴트) 위치 매개변수들; (ii) 6DoF 공간 설명 매개변수들(객체 좌표들); (iii) (선택적) 객체 방향성 매개변수들; (iv) (선택적) VR/AR/MR 환경 매개변수들; 및/또는 (v) (선택적) 거리 감쇠 매개변수들, 폐색 매개변수들, 잔향 매개변수들 등 중 적어도 하나를 포함할 수 있다.

[0141] 본 개시는 다음의 이점을 제공할 수 있다:

[0142] • 3DoF 오디오 디코딩 및 렌더링(예를 들어, MPEG-H 3DA 디코딩 및 렌더링)에 대한 하위 호환성: 6DoF 오디오 렌더러(예를 들어, MPEG-I 오디오 렌더러) 출력은 사전 결정된 3DoF 위치(들)에 대한 3DoF 렌더링 엔진(예를 들어, MPEG-H 3DA 렌더링 엔진)의 3DoF 렌더링 출력에 대응한다.

[0143] • 코딩 효율성: 이 접근법에서는, 레거시 3DoF 오디오 비트스트림 신택스(예를 들어, MPEG-H 3DA 비트스트림 신택스) 구조가 효율적으로 재사용될 수 있다.

[0144] • 사전 결정된 (3DoF) 위치(들)에서 오디오 품질 제어: 인코더가 어떤 임의의 위치(들) 및 대응하는 6DoF 공간에 대한 최상의 지각적 오디오 품질이 명시적으로 인코더에 의해 보장될 수 있다.

[0145] 본 개시의 예시적인 양상들은, MPEG 표준(예를 들어, MPEG-I 표준) 비트스트림과 호환되는 포맷의 다음의 시그널링:

[0146] • 확장 컨테이너 메커니즘(예를 들어, MPEG-H 3DA BS)을 통한 암시적 3DoF 오디오 시스템(예를 들어, MPEG-H 3DA) 호환성 시그널링 - 이는 6DoF 오디오(예를 들어, MPEG-I 오디오 호환) 처리 알고리즘으로 원래의 오디오 객체 신호들을 복구할 수 있게 함 -

[0147] • 원래의 오디오 객체 신호들의 근사를 위한 데이터를 설명하는 매개변수화

[0148] 에 관한 것일 수 있다.

[0149] 6DoF 오디오 렌더러는 MPEG 호환 시스템(예를 들어, MPEG-I 오디오 시스템)에서 원래의 오디오 객체 신호들을 복구하는 방법을 지정할 수 있다.

[0150] 이 제안된 개념은:

[0151] • 근사 함수(즉, $A(x)$)의 정의와 관련하여 일반적(generic)이고;

[0152] • 임의적으로 복잡할 수 있지만, 디코더 측에서 대응하는 근사가 존재해야 하고(예를 들어, $\exists A^{-1}$);

[0153] • 대략 수학적으로 (예를 들어, 알고리즘적으로 안정한 등의) "잘-정의된" 것일 수 있고;

[0154] • 근사 함수(예를 들어, $A(x)$)의 유형 면에서 일반적이고;

[0155] • 근사 함수는 다음의 근사 유형 또는 이런 접근법들의 임의의 조합에 기반할 수 있고(비트 레이트 소비 증가의 순으로 나열됨):

[0156] - 신호 x_{3DA}에 적용되는 매개변수화된 오디오 효과(들)(예를 들어, 매개변수적으로 제어된 레벨, 잔향, 반사, 폐색 등)

[0157] - 매개변수적으로 코딩된 수정(들)(예를 들어, 전송된 신호 x_{3DA}에 대한 시간/주파수 가변 수정 이득들)

[0158] - 신호 코딩된 수정(들)(예를 들어, 잔차 파형($x - x_{3DA}$)을 근사한 코딩된 신호들); 그리고

[0159] • 일반적 음장 및 음원 표현(및 그들의 조합): 객체들, 채널들, FOA, HOA으로 확장 가능하고 적용 가능하다.

[0160] 도 6a는 본 개시의 예시적인 양상들에 따른 예시적인 데이터 표현 및/또는 비트스트림 구조를 개략적으로 도시한다. 데이터 표현 및/또는 비트스트림 구조는 MPEG 표준(예를 들어, MPEG-H 또는 MPEG-I)과 호환되는 장치 또는 시스템을 통해(예를 들어, 소프트웨어, 하드웨어 또는 클라우드를 통해) 인코딩되었을 수 있다.

[0161] 비트스트림 BS는 예시적으로 3DoF 인코딩된 오디오 데이터를 포함하는 제1 비트스트림 부분(302)을 (예를 들어, 비트스트림의 주요 부분 또는 코어 부분에) 포함한다. 바람직하게는, 비트스트림 BS의 비트스트림 신택스는, 예를 들어, MPEG-H 3DA 비트스트림 신택스와 같은, 3DoF 오디오 렌더링의 BS 신택스와 호환되거나, 이를 준수한다. 3DoF 인코딩된 오디오 데이터는 비트스트림 BS의 하나 이상의 패킷에 페이로드로서 포함될 수 있다.

[0162] 예를 들어, 위의 도 3과 관련하여 이전에 설명된 바와 같이, 3DoF 인코딩된 오디오 데이터는 (예를 들어, 디폴트 3DoF 위치 주위의 구체 상에 있는) 하나 이상의 오디오 객체의 오디오 객체 신호들을 포함할 수 있다. 방향성 오디오 객체들의 경우, 3DoF 인코딩된 오디오 데이터는 선택적으로 객체 방향들을 더 포함할 수 있고, 및/또는 선택적으로 (예를 들어, 이득 및/또는 하나 이상의 감쇠 매개변수의 사용에 의해) 객체 거리들을 더 나타낼 수 있다.

[0163] 예시적으로, BS는 예시적으로 6DoF 오디오 인코딩을 위한 6DoF 메타데이터를 포함하는 제2 비트스트림 부분(303)을 (예를 들어, 비트스트림의 확장 부분 또는 메타데이터 부분에) 포함한다. 바람직하게는, 비트스트림 BS의 비트스트림 신택스는, 예를 들어, MPEG-H 3DA 비트스트림 신택스와 같은, 3DoF 오디오 렌더링의 BS 신택스와 호환되거나, 이를 준수한다. 6DoF 메타데이터는, 비트스트림 BS의 하나 이상의 패킷에(예를 들어, MPEG-H 3DA 비트스트림 구조에 의해 이미 제공된 하나 이상의 확장 컨테이너에) 확장 메타데이터로서 포함될 수 있다.

[0164] 예를 들어, 위의 도 3과 관련하여 이전에 설명된 바와 같이, 6DoF 메타데이터는 하나 이상의 3DoF(디폴트) 위치의 위치 데이터(예를 들어, 좌표(들)), 추가로 선택적으로 6DoF 공간 설명(예를 들어, 객체 좌표들), 추가로 선택적으로 객체 방향성, 추가로 선택적으로 VR 환경을 설명 및/또는 매개변수화하는 메타데이터, 및/또는 추가로 선택적으로 감쇠, 폐색 및/또는 잔향 등에 대한 매개변수화 정보 및/또는 매개변수들을 포함한다.

[0165] 도 6b는 본 개시의 예시적인 양상들에 따른 도 6a의 데이터 표현 및/또는 비트스트림 구조에 기반하는 예시적인 3DoF 오디오 렌더링을 개략적으로 도시한다. 도 6a에서와 같이, 데이터 표현 및/또는 비트스트림 구조는 MPEG 표준(예를 들어, MPEG-H 또는 MPEG-I)과 호환되는 장치 또는 시스템을 통해(예를 들어, 소프트웨어, 하드웨어 또는 클라우드를 통해) 인코딩되었을 수 있다.

[0166] 구체적으로, 3DoF 오디오 렌더링은, 6DoF 메타데이터를 폐기하여, 제1 비트스트림 부분(302)으로부터 획득된 3DoF 인코딩된 오디오 데이터에만 기반하여 3DoF 오디오 렌더링을 수행할 수 있는, 3DoF 오디오 렌더러에 의해 달성될 수 있다는 것이 도 6b에 예시적으로 도시되어 있다. 즉, MPEG-H 3DA 하위 호환성의 경우에, MPEG-H 3DA 렌더러는, 제1 비트스트림 부분(302)으로부터 획득된 3DoF 인코딩된 오디오 데이터에만 기반하여 효율적인 정규 MPEG-H 3DA 3DoF(또는 3DoF+) 오디오 렌더링을 수행하도록, 비트스트림의 확장 부분(예를 들어, 확장 컨테이너

(들))에 있는 6DoF 메타데이터를 효율적이며 안정적으로 방치/폐기할 수 있다.

[0167] 도 6c는 본 개시의 예시적인 양상들에 따른 도 6a의 데이터 표현 및/또는 비트스트림 구조에 기반하는 예시적인 6DoF 오디오 렌더링을 개략적으로 도시한다. 도 6a에서와 같이, 데이터 표현 및/또는 비트스트림 구조는 MPEG 표준(예를 들어, MPEG-H 또는 MPEG-I)과 호환되는 장치 또는 시스템을 통해(예를 들어, 소프트웨어, 하드웨어 또는 클라우드를 통해) 인코딩되었을 수 있다.

[0168] 구체적으로, 6DoF 오디오 렌더링은, 제1 비트스트림 부분(302)으로부터 획득된 3DoF 인코딩된 오디오 데이터 및 제2 비트스트림 부분(303)으로부터 획득된 6DoF 메타데이터에 기반하여 6DoF 오디오 렌더링을 수행하기 위해 제2 비트스트림 부분(303)으로부터 획득된 6DoF 메타데이터와 함께 제1 비트스트림 부분(302)으로부터 획득된 3DoF 인코딩된 오디오 데이터를 사용하는, (예를 들어 MPEG-I 또는 이후 표준에 따른) 새로운(novel) 6DoF 오디오 렌더러에 의해 달성될 수 있다는 것이 도 6c에 예시적으로 도시되어 있다.

[0169] 따라서, 비트스트림에서 용장성 없이 또는 적어도 감소된 상태로, 동일한 비트스트림이, 간단하고 유익한 하위 호환성을 허용하는, 3DoF 오디오 렌더링을 위한 레거시 3DoF 오디오 렌더러에 의해 사용될 수 있고, 6DoF 오디오 렌더링을 위한 새로운 6DoF 오디오 렌더러에 의해 사용될 수 있다.

[0170] 도 7a는 본 개시의 예시적인 양상들에 따른 3DoF 오디오 신호 데이터에 기반하는 6DoF 오디오 인코딩 변환 A를 개략적으로 도시한다. 변환(및 임의의 역변환)은 MPEG 표준(예를 들어, MPEG-H 또는 MPEG-I)과 호환되는 방법, 프로세스, 장치 또는 시스템에 따라 (예를 들어, 소프트웨어, 하드웨어 또는 클라우드를 통해) 수행될 수 있다.

[0171] 예시적으로, 위의 도 2 및 도 3과 마찬가지로, 도 7a는, 예시적으로 (벽들 (203) 뒤에 위치될 수 있거나 그의 사운드 신호가 감쇠, 잔향 및/또는 폐색 효과를 초래할 수 있는 다른 구조에 의해 방해받을(obstructed) 수 있는) 복수의 오디오 소스(207)를 포함하는, 방의 예시적인 평면도(202)를 도시한다.

[0172] 3DoF 오디오 렌더링을 위해, 복수의 오디오 소스(207)의 오디오 신호들 x는 디폴트 3DoF 위치(206)(예를 들어, 3DoF 음장에서 청취자 위치) 주위의 구체 S 상의 3DoF 오디오 신호들(오디오 객체들)을 획득하도록 변환된다. 위와 같이, 3DoF 오디오 신호들은 x_{3DA}로 지칭되고 다음과 같은 변환 함수 A를 사용하여 획득될 수 있다:

[0173] [수학식 6]

$$x_{3DA} = A(x)$$

[0174]

[0175] 위의 수학식에서, x는 음원(들)/객체 신호(들)를 나타내고, x_{3DA}는 디폴트 3DoF 위치(206)에서 동일한 음장을 생성하는 3DA를 위한 대응하는 가상 3DA 객체 신호들을 나타내고, A는 오디오 신호 x에 기반하여 오디오 신호들 x_{3DA}를 근사하는 변환 함수를 나타낸다. 역변환 함수 A^{-1}는, 이미 위에서 그리고 아래에서 더 논의되는 바와 같이, 6DoF 오디오 렌더링을 위한 음원 신호들을 복원/근사하는 데에 사용될 수 있다. $A\,A^{-1} = 1$ 및 $A^{-1}A = 1$ 또는 적어도 $A\,A^{-1} \approx 1$ 및 $A^{-1}\,A \approx 1$ 임에 유의하자.

[0176] 일반적인 방식으로, 변환 함수 A는 본 개시의 일부 예시적인 양상에서 디폴트 3DoF 위치(206)를 둘러싸는 구체 S 상에 오디오 신호들 x를 투사하거나 적어도 매핑하는 매핑/투사 함수로서 간주될 수 있다.

[0177] 3DoF 오디오 렌더링은 VR 환경(예컨대, 감쇠, 잔향, 폐색 효과 등을 초래할 수 있는 기존의 벽들(203) 등 또는 기타 구조)을 인식하지 못한다는 점에 더 유의해야 한다. 따라서, 변환 함수 A는 바람직하게는 이러한 VR 환경적 특성들에 기반하는 효과들을 포함할 수 있다.

[0178] 도 7b는 본 개시의 예시적인 양상들에 따라 3DoF 오디오 신호 데이터에 기반하여 6DoF 오디오 신호 데이터를 근사/복원하기 위한 6DoF 오디오 디코딩 변환 A^{-1}을 개략적으로 도시한다.

[0179] 역변환 함수 A^{-1} 및 위의 도 7a에서와 같이 획득된 근사된 3DoF 오디오 신호들 x_{3DA}를 사용함으로써, 원래의 오디오 소스(207)의 원래의 오디오 신호 x^*는 다음과 같이 복원/근사될 수 있다:

[0180] [수학식 7]

[0181] $$x^* = A^{-1}(x_{3DA})$$

[0182] 따라서, 도 7b의 오디오 객체(320)들의 오디오 신호 x^*는 원래의 소스들(207)의 오디오 신호들 x과 유사하거나 동일하게, 구체적으로 원래의 소스들(207)과 동일한 위치들에서 복원될 수 있다.

[0183] 도 7c는, 본 개시의 예시적인 양상들에 따른, 도 7b의 근사/복원된 6DoF 오디오 신호 데이터에 기반하는 예시적인 6DoF 오디오 렌더링을 개략적으로 도시한다.

[0184] 그 다음, 도 7b의 오디오 객체(320)들의 오디오 신호들 x^*는 6DoF 오디오 렌더링에 사용될 수 있고, 여기서 청취자의 위치는 또한 가변된다.

[0185] 청취자의 청취자 위치가 위치(206)(디폴트 3DoF 위치와 동일한 위치)에 있다고 가정될 때, 6DoF 오디오 렌더링은 오디오 신호들 x_{3DA}에 기반하는 3DoF 오디오 렌더링과 동일한 음장을 렌더링한다.

[0186] 따라서, 가정되고 있는 청취자 위치인 디폴트 3DoF 위치에서의 6DoF 렌더링 $F_{6DoF}(x^*)$는 3DoF 렌더링 $F_{3DoF}(x_{3DA})$와 동일(또는 적어도 대략 동일)하다.

[0187] 또한, 청취자 위치가 예를 들어 도 7c에서 위치 206'로 이동되면, 6DoF 오디오 렌더링에서 생성된 음장은 상이해지지만, 바람직하게는 원활하게 발생할 수 있다.

[0188] 또다른 예시로서, 제3 청취자 위치(206")가 가정될 수 있고, 6DoF 오디오 렌더링 시에 생성된 음장은 구체적으로 좌측 상단 오디오 신호에 대해 상이해지며, 이는 제3 청취자 위치(206")에 대해 벽(203)에 의해 방해받지 않는다. 바람직하게는, 이것은, 역함수 A^{-1}이(VR 환경 특성들과 같은 환경적 효과들 없이) 원래의 음원을 복원하기 때문에, 가능해진다.

[0189] 도 8은 본 개시의 예시적인 양상들에 따른 3DoF/6DoF 비트스트림 인코딩 방법의 예시적인 흐름도를 개략적으로 도시한다. 단계들의 순서는 제한되지 않으며 상황에 따라 변경될 수 있다는 점에 유의해야 한다. 또한, 방법의 일부 단계는 선택적이라는 점에 유의해야 한다. 예를 들어, 본 방법은 디코더, 오디오 디코더, 오디오/비디오 디코더 또는 디코더 시스템에 의해 실행될 수 있다.

[0190] 단계 S801에서, 본 방법은 (예를 들어, 디코더 측에서) 하나 이상의 오디오 소스의 원래의 오디오 신호(들) x를 수신한다.

[0191] 단계 S802에서, 본 방법은 (선택적으로) 환경 특성들(예를 들어, 방 형상, 벽들, 벽 사운드 반사 특성들, 객체들, 장애물들 등)을 결정하고, 및/또는 (감쇠, 이득, 폐색, 잔향 등의 효과들을 매개변수화한) 매개변수들을 결정한다.

[0192] 단계 S803에서, 본 방법은 (선택적으로) 예를 들어, 단계 S802의 결과에 기반하여 변환 함수 A의 매개변수화를 결정한다. 바람직하게는, 단계 S803은 매개변수화되거나 사전 설정된 변환 함수 A를 제공한다.

[0193] 단계 S804에서, 본 방법은 변환 함수 A에 기반하여 하나 이상의 오디오 소스의 원래의 오디오 신호(들) x를 대응하는 하나 이상의 근사된 3DoF 오디오 신호(들) x_{3DA}로 변환한다.

[0194] 단계 S805에서, 본 방법은 (하나 이상의 3DoF 위치, VR 환경 정보 및/또는 감쇠, 이득, 폐색, 잔향 등과 같은 환경적 효과들의 매개변수화 및 매개변수들을 포함할 수 있는) 6DoF 메타데이터를 결정한다.

[0195] 단계 S806에서, 본 방법은 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)에 3DoF 오디오 신호(들) x_{3DA}를 포함(내장)한다.

[0196] 단계 S807에서, 본 방법은 제2 비트스트림 부분(또는 다수의 제2 비트스트림 부분)에 6DoF 메타데이터를 포함(내장)한다.

[0197] 그 다음, 단계 S808에서, 본 방법은, 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)에 3DoF 오디오 신호(들) x_{3DA}를, 그리고 제2 비트스트림 부분(또는 다수의 제2 비트스트림 부분)에 6DoF 메타데이터를 포함하는 인코딩된 비트스트림을 제공하기 위해, 제1 및 제2 비트스트림 부분에 기반하여 비트스트림을 계속해서 인코딩

한다.

[0198] 그 다음, 인코딩된 비트스트림은, 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)에서의 3DoF 오디오 신호(들) x_{3DA}에만 기반하는 3DoF 오디오 렌더링을 위한 3DoF 디코더/렌더러에 제공되거나, 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)에서의 3DoF 오디오 신호(들) x_{3DA} 및 제2 비트스트림 부분(또는 다수의 제2 비트스트림 부분)에서의 6DoF 메타데이터에 기반하는 6DoF오디오 렌더링을 위한 6DoF 디코더/렌더러에 제공될 수 있다.

[0199] 도 9는 본 개시의 예시적인 양상들에 따른 3DoF 및/또는 6DoF 오디오 렌더링의 방법들의 예시적인 흐름도를 개략적으로 도시한다. 단계들의 순서는 제한되지 않으며 상황에 따라 변경될 수 있다는 점에 유의해야 한다. 또한, 방법의 일부 단계는 선택적이라는 점에 유의해야 한다. 본 방법은, 예를 들어, 인코더, 렌더러, 오디오 인코더, 오디오 렌더러, 오디오/비디오 인코더 또는 인코더 시스템 또는 렌더러 시스템에 의해 실행될 수 있다.

[0200] 단계 S901에서, 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)에 3DoF 오디오 신호(들) x_{3DA}를, 그리고 제2 비트스트림 부분(또는 다수의 제2 비트스트림 부분)에 6DoF 메타데이터를 포함하는 인코딩된 비트스트림이 수신된다.

[0201] 단계 S902에서, 3DoF 오디오 신호(들) x_{3DA}는 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)으로부터 획득된다. 이는 3DoF 디코더/렌더러와 또한 6DoF 디코더/렌더러에 의해 수행될 수 있다.

[0202] 디코더/렌더러가 3DoF 오디오 렌더링을 위한 레거시 장치(또는 3DoF 오디오 렌더링 모드로 전환되는 신규 3DoF/6DoF 디코더/렌더러)이면, 본 방법은 6DoF 메타데이터가 폐기/방치되는 단계 S903으로 진행하고, 그런 다음, 3DoF 오디오 렌더링 동작을 진행하여 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)으로부터 획득된 3DoF 오디오 신호(들) x_{3DA}에 기반하여 3DoF 오디오를 렌더링한다.

[0203] 즉, 하위 호환성이 유리하게 보장된다.

[0204] 한편, 디코더/렌더러가 6DoF 오디오 렌더링을 위한 것(예를 들어, 신규 6DoF 디코더/렌더러 또는 6DoF 오디오 렌더링 모드로 전환되는 3DoF/6DoF 디코더/렌더러)이면, 본 방법은 단계 S905로 진행하여 제2 비트스트림 부분(들)로부터 6DoF 메타데이터를 획득한다.

[0205] 단계 S906에서, 본 방법은, 제2 비트스트림 부분(또는 다수의 제2 비트스트림 부분)으로부터 획득된 6DoF 메타데이터 및 역변환 함수 A^{-1}에 기반하여, 제1 비트스트림 부분(또는 다수의 제1 비트스트림 부분)으로부터 획득된 3DoF 오디오 신호(들) x_{3DA}로부터 오디오 객체들/소스들의 오디오 신호들 x^*를 근사/복원한다.

[0206] 그 다음, 단계 S907에서, 본 방법은 오디오 객체들/소스들의 근사/복원된 오디오 신호들 x^*에 기반하여, 그리고 (VR 환경 내에서 가변적일 수 있는) 청취자 위치에 기반하여, 6DoF 오디오 렌더링을 수행하는 것으로 진행한다.

[0207] 위의 예시적인 양상들에서, 3D 오디오 인코딩 및/또는 3D 오디오 렌더링을 위한 효율적이며 신뢰할 수 있는 방법, 장치 및 데이터 표현 및/또는 비트스트림 구조가 제공될 수 있고, 이는, 유익하게도, 예를 들어, MPEG-H 3DA 표준에 따른 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 허용한다. 구체적으로, 바람직하게는, 예를 들어 MPEG-H 3DA 표준에 따른 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 허용하는 3D 오디오 인코딩 및/또는 3D 오디오 렌더링을 위한 데이터 표현 및/또는 비트스트림 구조와, 예를 들어 MPEG-H 3DA 표준에 따른 3DoF 오디오 렌더링에 대한 하위 호환성을 갖는, 효율적인 6DoF 오디오 인코딩 및/또는 렌더링을 위한 대응하는 인코딩 및/또는 렌더링 장치를 제공하는 것이 가능하다.

[0208] 본 명세서에 설명된 방법 및 시스템은 소프트웨어, 펌웨어 및/또는 하드웨어로 구현될 수 있다. 특정 구성 요소는 디지털 신호 프로세서 또는 마이크로프로세서 상에서 실행되는 소프트웨어로 구현될 수 있다. 기타 구성 요소는 하드웨어 및/또는 응용 주문형 집적 회로(application specific integrated circuits)로 구현될 수 있다. 설명된 방법 및 시스템에서 접하게 되는 신호들은 랜덤 액세스 메모리 또는 광학 저장 매체와 같은 매체에 저장될 수 있다. 이들은 무선 네트워크, 위성 네트워크, 와이어리스 네트워크 또는 유선 네트워크(예를 들어, 인터넷)와 같은 네트워크를 통해 전송될 수 있다. 본 명세서에 설명된 방법 및 시스템을 사용하는 전형적인 디바이스는 오디오 신호들을 저장 및/또는 렌더링하는 데에 사용되는 휴대용 전자 디바이스 또는 기타 고객 장비

(consumer equipment)이다.

[0209] 본 개시에 따른 방법 및 장치의 예시적인 구현은 청구항들이 아닌 다음의 열거된 예시적인 실시예(EEE)로부터 명백해질 것이다.

[0210] EEE1은 예시적으로 오디오 소스 신호들, 3DoF 관련 데이터 및 6DoF 관련 데이터를 포함하는 오디오를 인코딩하는 방법에 관한 것으로, 예를 들어, 오디오 소스 장치에 의해, 특히 인코더에서와 같이, 3DoF 데이터를 결정하기 위해 3DoF 위치(들)에서 원하는 음장을 근사하는 오디오 소스 신호들을 인코딩하는 단계; 및/또는 예를 들어, 오디오 소스 장치에 의해, 특히 인코더에서와 같이, 6DoF 메타데이터를 결정하기 위해 6DoF 관련 데이터를 인코딩하는 단계 - 메타데이터는 6DoF 렌더링을 위해 원래의 오디오 소스 신호를 근사하는 데에 사용될 수 있음 - 를 포함한다.

[0211] EEE2는 예시적으로 EEE1의 방법에 관한 것으로서, 3DoF 데이터는 객체 오디오 신호들, 객체 방향들 및 객체 거리들 중 적어도 하나에 관한 것이다.

[0212] EEE3는 예시적으로 EEE1 또는 EEE2의 방법에 관한 것으로서, 6DoF 데이터는 3DoF (디폴트) 위치 매개변수들, 6DoF 공간 설명(객체 좌표들) 매개변수들, 객체 방향성 매개변수들, VR 환경 매개변수들, 거리 감쇠 매개변수들, 폐색 매개변수들 및 잔향 매개변수들 중 적어도 하나에 관한 것이다.

[0213] EEE4는 예시적으로 데이터, 특히 3DoF 및 6DoF 렌더링 가능한 오디오 데이터를 전송하는 방법에 관한 것으로서, 이 방법은, 예를 들어, 3DoF 오디오 시스템에 의해 디코딩될 때, 예를 들어, 오디오 비트스트림 신택스에서, 바람직하게는 3DoF 위치(들)에서 원하는 음장을 근사할 수 있는 오디오 소스 신호들을 전송하는 단계; 및/또는 예를 들어, 오디오 비트스트림 신택스의 확장 부분에서, 6DoF 렌더링을 위해 원래의 오디오 소스 신호들을 근사 및/또는 복원하기 위한 6DoF 관련 메타데이터를 전송하는 단계 - 6DoF 관련 메타데이터는 파라메트릭 데이터 및/또는 신호 데이터일 수 있음 - 를 포함한다.

[0214] EEE5는 예시적으로 EEE4의 방법에 관한 것으로서, 예를 들어 3DoF 메타데이터 및/또는 6DoF 메타데이터를 포함하는 오디오 비트스트림 신택스는 적어도 MPEG-H 오디오 표준의 버전에 대해 불만(complaint with)이다.

[0215] EEE6는 예시적으로 비트스트림을 생성하는 방법에 관한 것으로서, 이 방법은, 3DoF 위치(들)에서 원하는 음장을 근사하는 오디오 소스 신호들에 기반하는 3DoF 메타데이터를 결정하는 단계; 6DoF 관련 메타데이터를 결정하는 단계 - 메타데이터는 6DoF 렌더링을 위해 원래의 오디오 소스 신호들을 근사하는 데에 사용될 수 있음 -; 및/또는 오디오 소스 신호 및 6DoF 관련 메타데이터를 비트스트림에 삽입하는 단계를 포함한다.

[0216] EEE7은 예시적으로 오디오 렌더링을 위한 방법에 관한 것으로서, 상기 방법은,

[0217] 3DoF 위치(들)에서 원래의 오디오 신호들 x의 근사된 오디오 신호들 x^*의 6DoF 메타데이터의 전처리를 포함하고, 6DoF 렌더링은 3DoF 위치(들)에서 원하는 음장을 근사하는 3DoF 렌더링을 위해 전송된 오디오 소스 신호들 x_{3DA}의 3DoF 렌더링과 동일한 출력을 제공할 수 있다.

[0218] EEE8은 예시적으로 EEE7의 방법에 관한 것으로서, 오디오 렌더링은:

[0219] $$F_{6DoF}(x^*) \approx F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}(x)\ for\ 3DoF$$

[0220] 에 기반하여 결정되고, 여기서 $F_{6DoF}(x^*)$는 6DoF 청취자 위치(들)에 대한 오디오 렌더링 함수에 관한 것이고, $F_{3DoF}(x_{3DA})$는 3DoF 청취자 위치(들)에 대한 오디오 렌더링 함수들에 관한 것이고, x_{3DA}는 특정 3DoF 위치(들)에 대한 VR 환경의 효과들을 포함하는 오디오 신호들이고, x^*는 근사된 오디오 신호들에 관한 것이다.

[0221] EEE9는 예시적으로 EEE8의 방법에 관한 것으로서, 원래의 오디오 신호들 x의 근사된 오디오 신호들 x^*는:

[0222] $$x^* := A^{-1}(x_{3DA})$$

[0223] 에 기반하고, 여기서 A^{-1}은 근사 함수 A의 역(inverse)에 관한 것이다.

[0224] EEE10은 예시적으로 EEE8 또는 EEE9의 방법에 관한 것으로서, 근사 방법 A를 사용하여 원래의 오디오 소스 신호들 x의 근사된 오디오 신호들 x^*를 획득하기 위해 사용되는 메타데이터는,

[0225] $$x_{3DA} := A(x), \qquad \left\| F_{3DoF}(x_{3DA}) - F_{6DoF}\ (x)\ for\ 3DoF \right\| \rightarrow min$$

[0226] 에 기반하여 정의되며, 여기서 메타데이터의 양은 원래의 오디오 소스 신호들 x를 전송하는 데에 필요한 오디오 데이터의 양보다 적고,

[0227] 여기서 오디오 렌더링은:

[0228] $$F_{6DoF}(x^*) \approx F_{3DoF}(x_{3DA}) \rightarrow F_{6DoF}\ (x)\ for\ 3DoF$$

[0229] 에 기반하여 결정되며, 여기서 $F_{6DoF}(x^*)$는 6DoF 청취자 위치(들)에 대한 오디오 렌더링 함수에 관한 것이고, $F_{3DoF}(x_{3DA})$는 3DoF 청취자 위치(들)에 대한 오디오 렌더링 함수들에 관한 것이고, x_{3DA}는 특정 3DoF 위치(들)에 대한 VR 환경의 효과들을 포함하는 오디오 신호들이고, x^*는 근사된 오디오 신호들에 관한 것이다.

[0230] 본 개시의 예시적인 양상들 및 실시예들은 하드웨어, 펌웨어, 또는 소프트웨어, 또는 이 둘의 조합으로(예를 들어, 프로그래밍 가능한 로직 어레이로서) 구현될 수 있다. 달리 지정되지 않는 한, 본 개시의 일부로서 포함된 알고리즘 또는 프로세스는 본질적으로 임의의 특정 컴퓨터 또는 기타 장치에 관련되지 않는다. 특히, 다양한 범용 기계가 본 명세서의 교시에 따라 기입된 프로그램과 함께 사용될 수 있거나, 요구된 방법의 단계들을 수행하기 위해 더 전용 장치(specialized apparatus)(예를 들어, 집적 회로)를 구성하는 것이 더 편리할 수 있다. 따라서, 본 개시는 각각 적어도 하나의 프로세서, (휘발성 및 비휘발성 메모리 및/또는 저장 요소들을 포함하는) 적어도 하나의 데이터 저장 시스템, 적어도 하나의 입력 디바이스 또는 포트, 및 적어도 하나의 출력 디바이스 또는 포트를 포함하는 하나 이상의 프로그램 가능한 컴퓨터 시스템(예를 들어, 도면의 임의의 요소의 구현물)상에서 실행되는 하나 이상의 컴퓨터 프로그램으로 구현될 수 있다. 프로그램 코드는 본 명세서에 설명된 기능들을 수행하고 출력 정보를 생성하기 위해 입력 데이터에 적용된다. 출력 정보는 알려진 방식으로 하나 이상의 출력 디바이스에 적용된다.

[0231] 각각의 이러한 프로그램은 컴퓨터 시스템과 통신하기 위해 (기계, 어셈블리 또는 고수준의 절차적, 논리적 또는 객체 지향된 프로그래밍 언어를 포함하는) 임의의 원하는 컴퓨터 언어로 구현될 수 있다. 임의의 경우에, 언어는 컴파일되거나 해석된 언어일 수 있다.

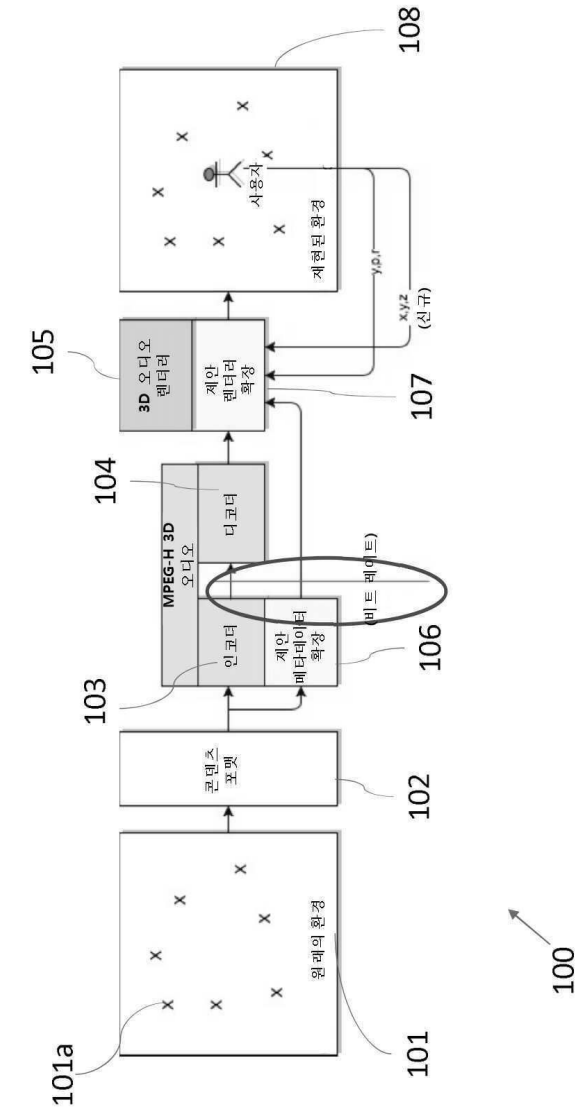
[0232] 예를 들어, 컴퓨터 소프트웨어 명령어 시퀀스들에 의해 구현될 때, 본 개시의 다양한 기능 및 단계는 적합한 디지털 신호 처리 하드웨어에서 실행되는 다중 스레드의 소프트웨어 명령어 시퀀스들에 의해 구현될 수 있고, 이 경우, 실시예들의 다양한 디바이스, 단계 및 기능이 소프트웨어 명령어들의 일부에 대응할 수 있다.

[0233] 각각의 이러한 컴퓨터 프로그램은, 본 명세서에 설명된 절차들을 수행하기 위해 저장 매체 또는 디바이스가 컴퓨터 시스템에 의해 판독될 때, 컴퓨터를 구성하고 동작시키기 위하여, 바람직하게는, 범용 또는 전용 프로그래밍 가능한 컴퓨터에 의해 판독 가능한 저장 매체 또는 디바이스(예를 들어, 고체 상태 메모리 또는 매체, 또는 자기 또는 광학 매체)에 저장되거나 다운로딩될 수 있다. 본 독창적인 시스템은 또한, 컴퓨터 프로그램으로 구성되는(즉, 이를 저장하는) 컴퓨터 판독 가능한 저장 매체로서 구현될 수 있고, 이렇게 구성된 저장 매체는 컴퓨터 시스템으로 하여금 본 명세서에 설명된 기능을 수행하기 위해 특정한 사전 정의된 방식으로 동작하게 한다.

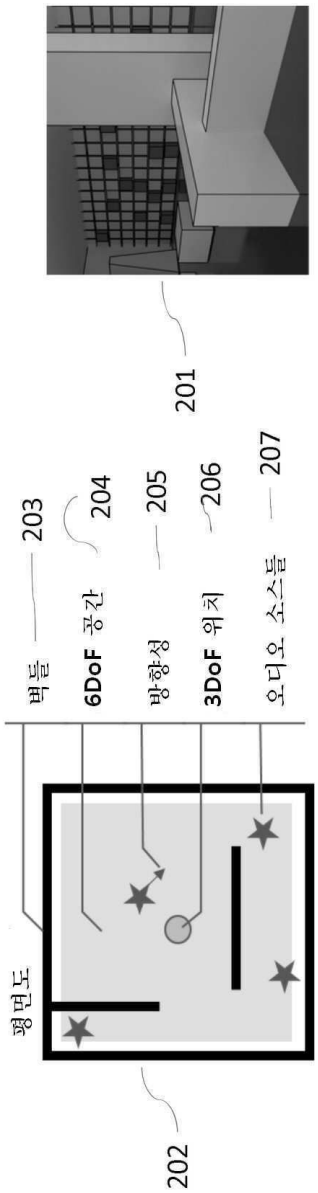
[0234] 본 개시의 발명의 다수의 예시적인 양상 및 예시적인 실시예에 대하여 위에서 설명했다. 그럼에도 불구하고, 본 개시의 사상 및 범위를 벗어남 없이 다양한 수정이 이루어질 수 있음을 이해할 것이다. 위의 교시를 고려하여 본 발명의 수많은 수정 및 변형이 가능하다. 첨부된 청구항들의 범위 내에서, 본 개시의 발명은 본 명세서에 구체적으로 설명된 것과 다르게 실시될 수 있다는 점을 이해해야 한다.

도면

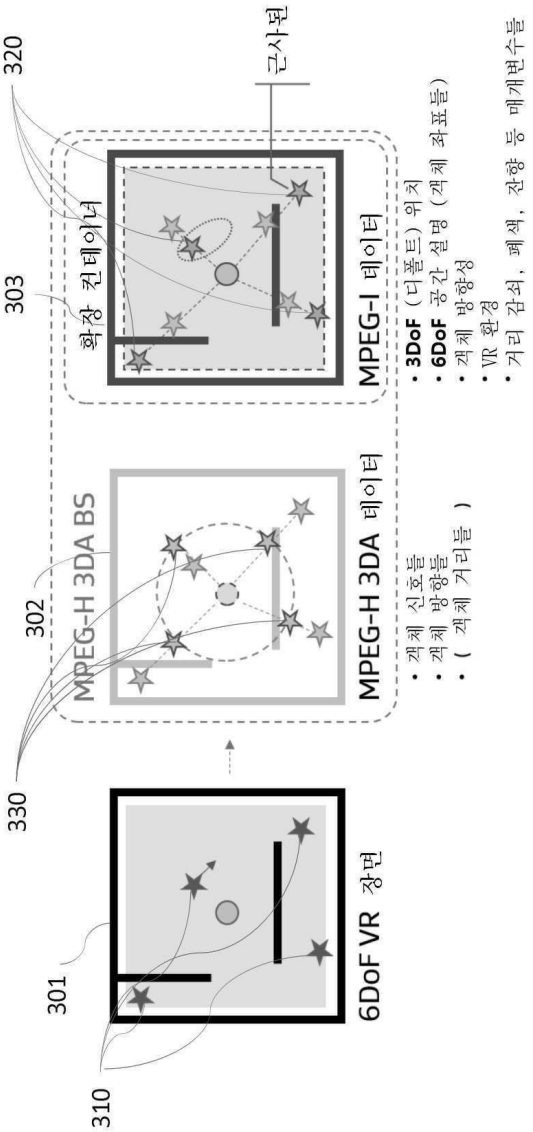
도면1


101a
원래의 환경
101
콘텐츠 포맷
102
인코더
103
MPEG-H 3D 오디오
디코더
104
3D 오디오 렌더러
105
제안 렌더러 확장
107
제안 메타데이터 확장
106
비트 레이트)
x,y,z (신규)
y,p,r
사용자
재현된 환경
108
100

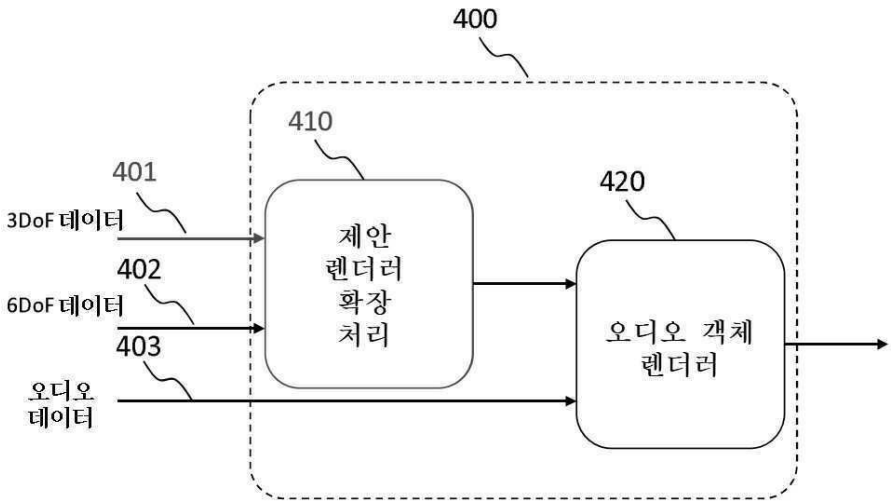
도면2



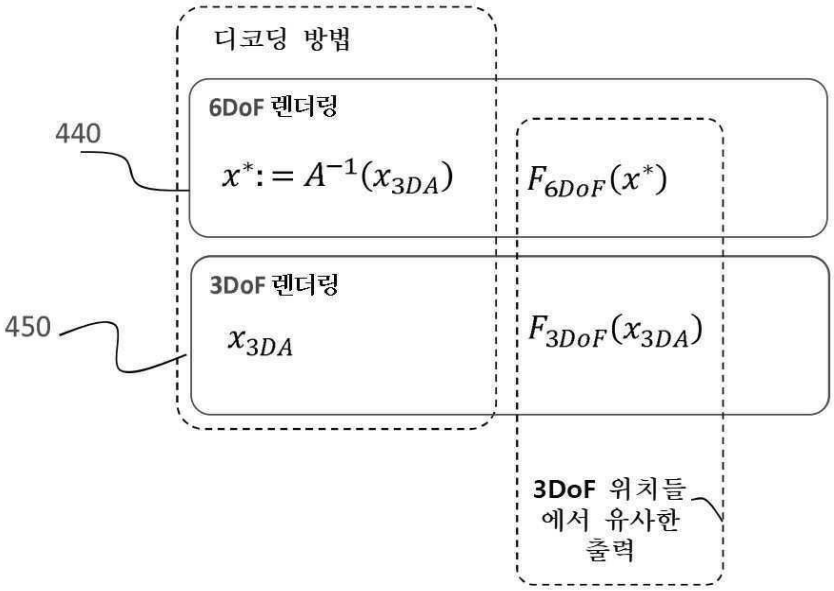
도면3


320
303
302
330
301
310
6DoF VR 장면
MPEG-H 3DA 데이터
MPEG-H 3DA BS
MPEG-I 데이터
확장 컨테이너
근사된
• 객체 신호들
• 객체 방향들
• (객체 거리들)
• 3DoF (디폴트) 위치
• 6DoF 공간 설명 (객체 좌표들)
• 객체 방향성
• VR 환경
• 거리 감쇠, 페색, 잔향 등 매개변수들

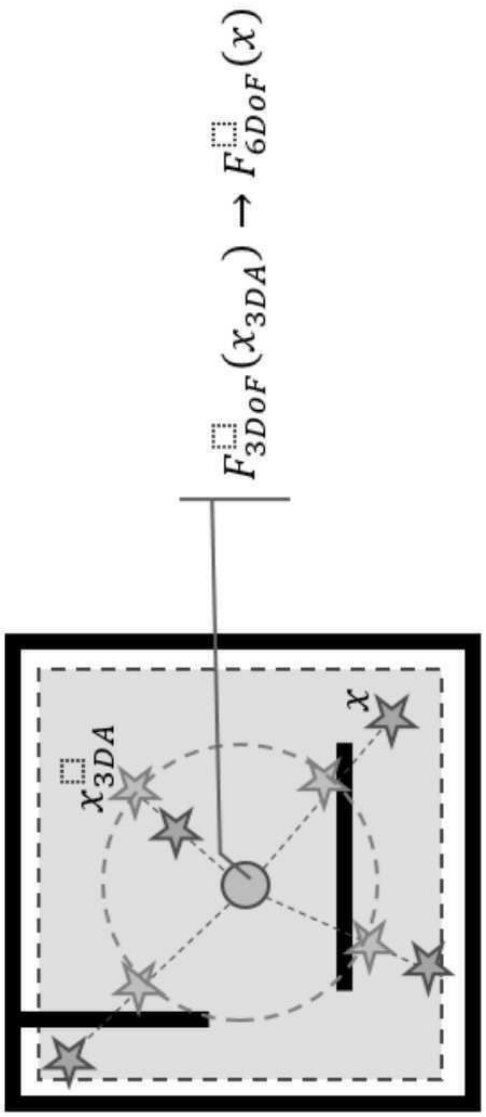
도면4a


400
410
401
3DoF 데이터
402
6DoF 데이터
403
오디오
데이터
제안
렌더러
확장
처리
420
오디오 객체
렌더러

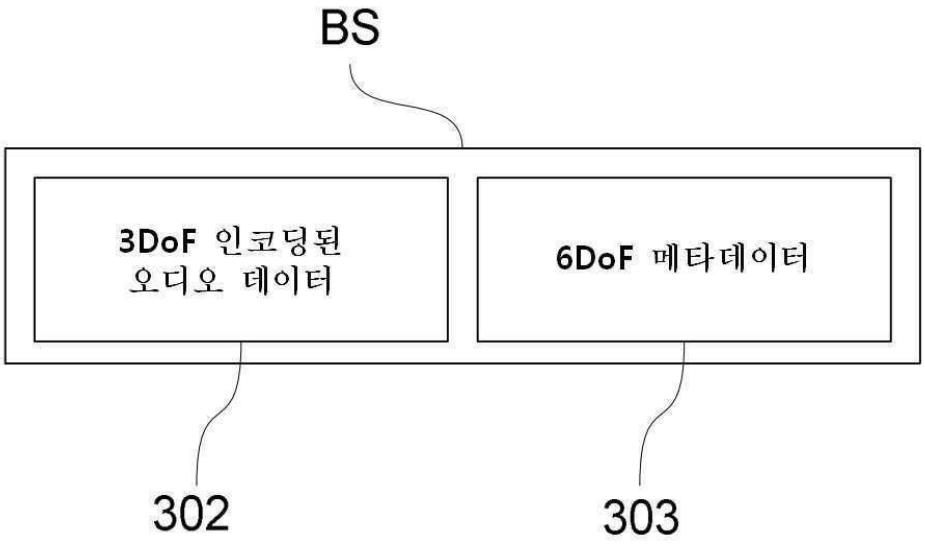
도면4b


디코딩 방법
6DoF 렌더링
440
$x^*: = A^{-1}(x_{3DA})$
$F_{6DoF}(x^*)$
3DoF 렌더링
450
x_{3DA}
$F_{3DoF}(x_{3DA})$
3DoF 위치들
에서 유사한
출력

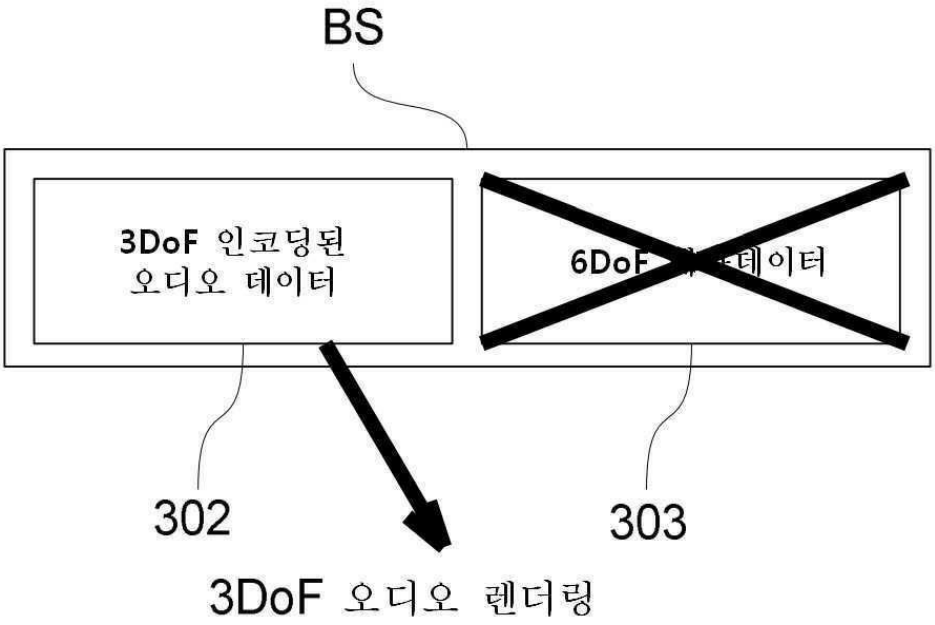
도면5



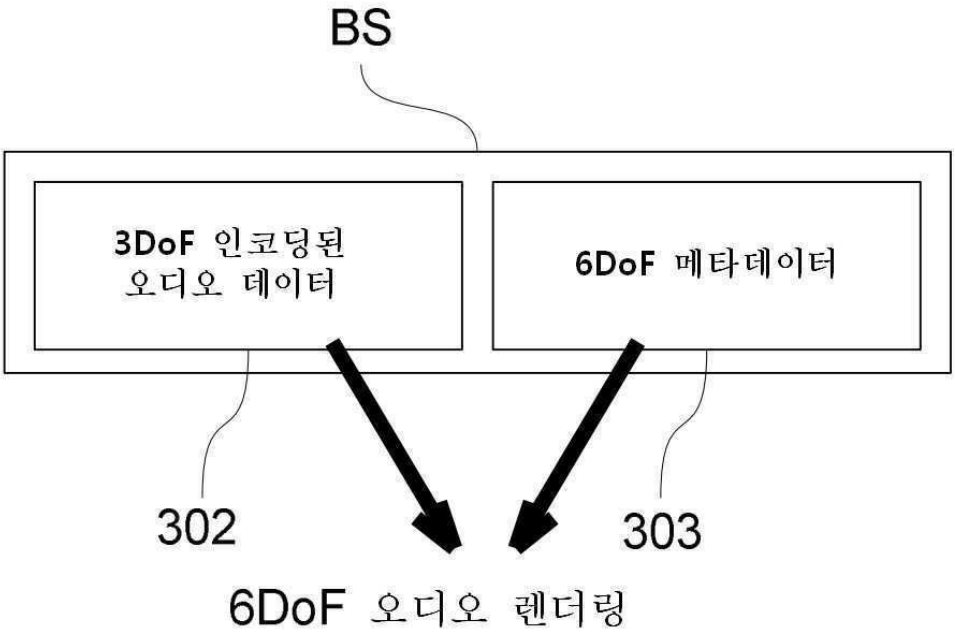
도면6a


BS
3DoF 인코딩된
오디오 데이터
6DoF 메타데이터
302
303

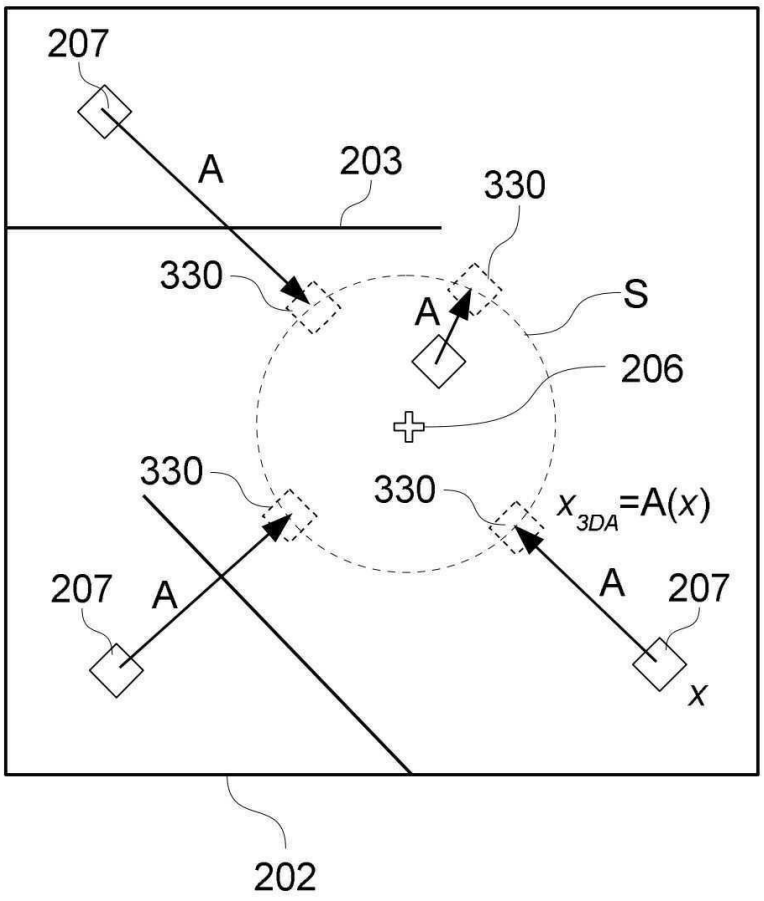
도면6b


BS
3DoF 인코딩된
오디오 데이터
6DoF
302
303
3DoF 오디오 렌더링

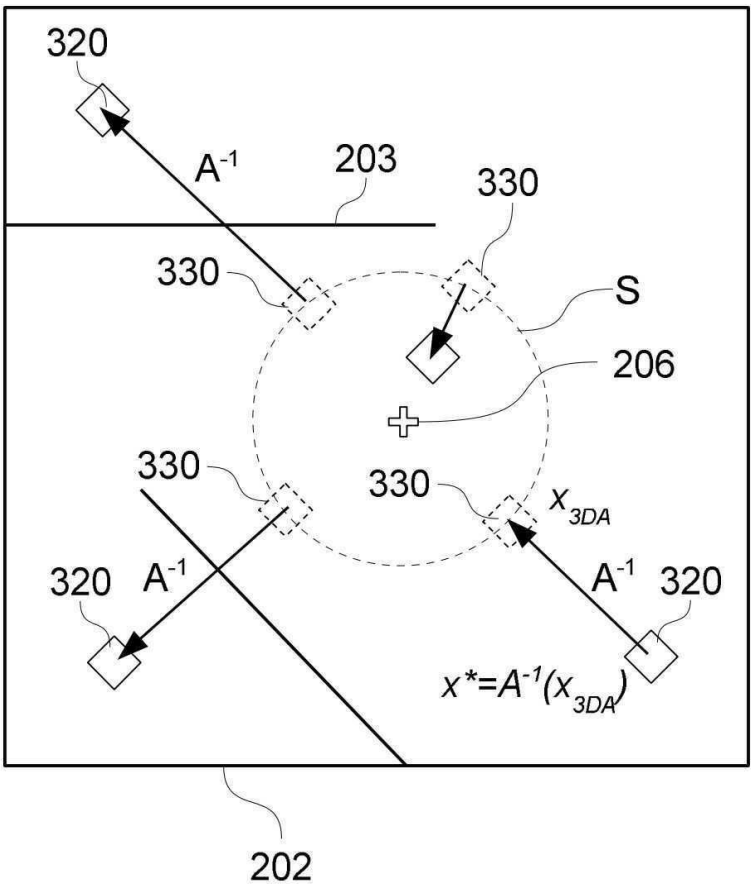
도면6c


BS
3DoF 인코딩된
오디오 데이터
6DoF 메타데이터
302
303
6DoF 오디오 렌더링

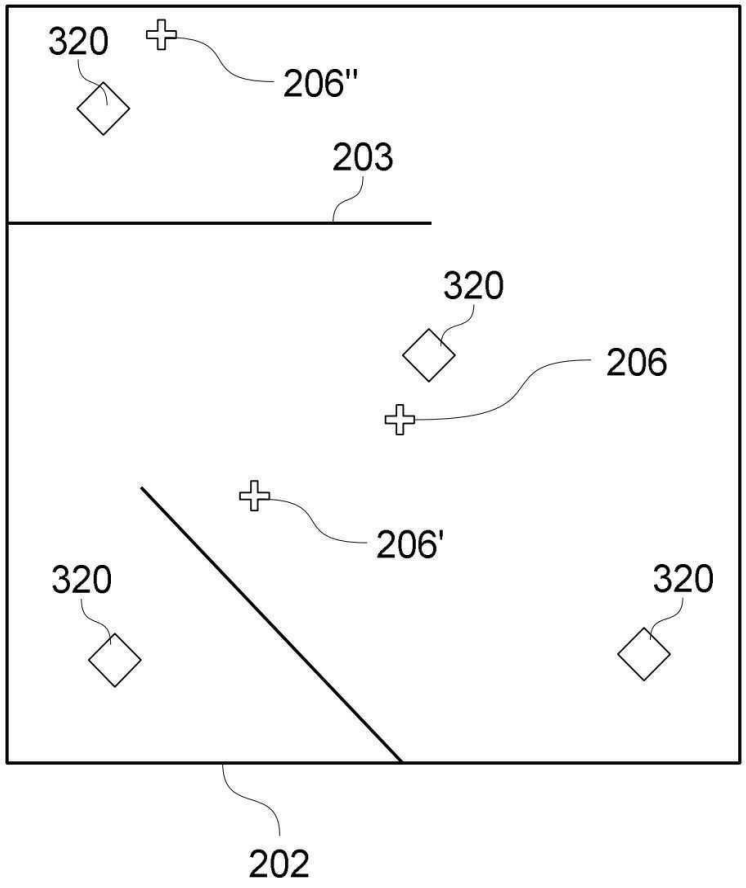
도면7a


207
A
203
330
330
A
S
206
330
330
x_{3DA}=A(x)
207
A
A
207
x
202

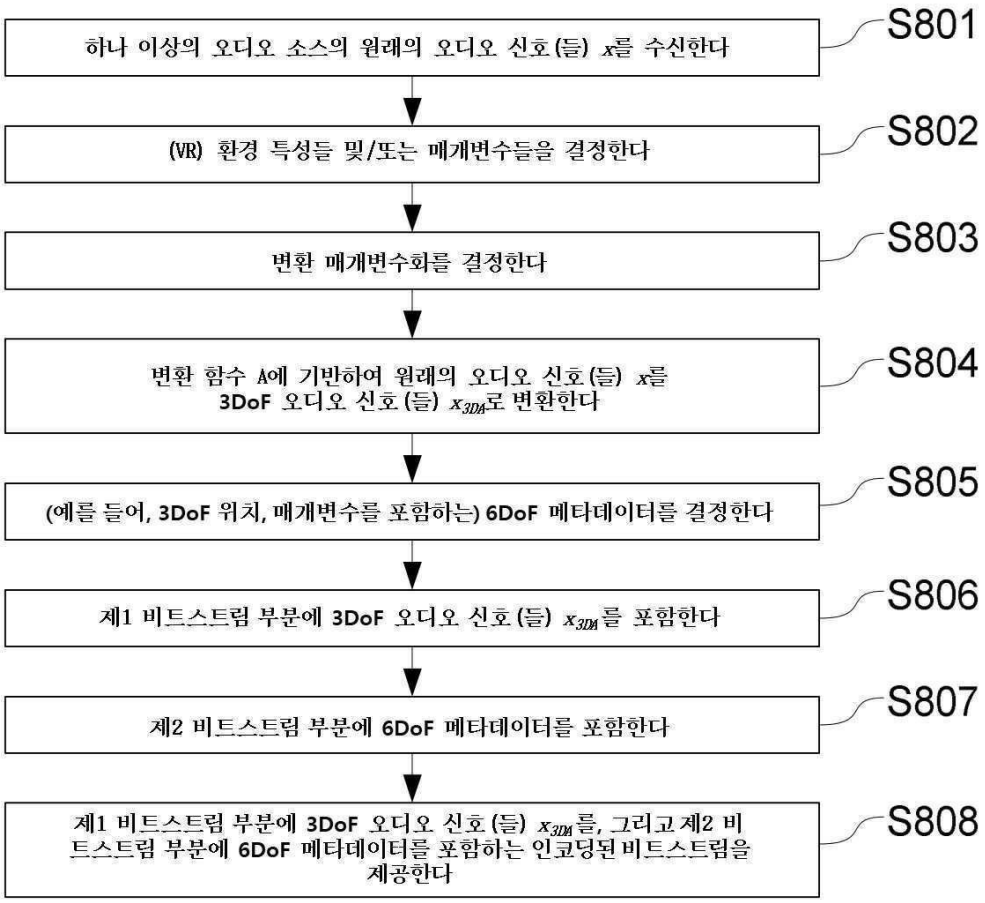
도면7b



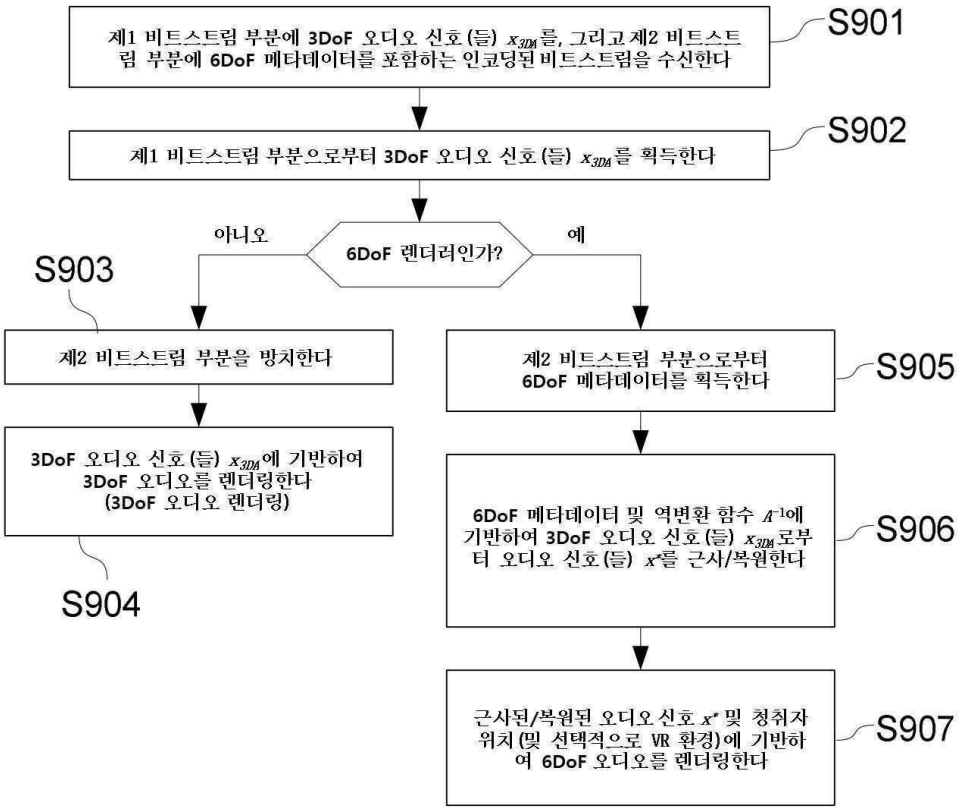
도면7c



도면8


하나 이상의 오디오 소스의 원래의 오디오 신호(들) x를 수신한다
S801
(VR) 환경 특성들 및/또는 매개변수들을 결정한다
S802
변환 매개변수화를 결정한다
S803
변환 함수 A에 기반하여 원래의 오디오 신호(들) x를 3DoF 오디오 신호(들) x_{3DA}로 변환한다
S804
(예를 들어, 3DoF 위치, 매개변수를 포함하는) 6DoF 메타데이터를 결정한다
S805
제1 비트스트림 부분에 3DoF 오디오 신호(들) x_{3DA}를 포함한다
S806
제2 비트스트림 부분에 6DoF 메타데이터를 포함한다
S807
제1 비트스트림 부분에 3DoF 오디오 신호(들) x_{3DA}를, 그리고 제2 비트스트림 부분에 6DoF 메타데이터를 포함하는 인코딩된 비트스트림을 제공한다
S808

도면9


제1 비트스트림 부분에 3DoF 오디오 신호(들) x_{3DA}를, 그리고 제2 비트스트림 부분에 6DoF 메타데이터를 포함하는 인코딩된 비트스트림을 수신한다
S901
제1 비트스트림 부분으로부터 3DoF 오디오 신호(들) x_{3DA}를 획득한다
S902
6DoF 렌더러인가?
아니오
예
S903
제2 비트스트림 부분을 방치한다
3DoF 오디오 신호(들) x_{3DA}에 기반하여 3DoF 오디오를 렌더링한다 (3DoF 오디오 렌더링)
S904
제2 비트스트림 부분으로부터 6DoF 메타데이터를 획득한다
S905
6DoF 메타데이터 및 역변환 함수 A^{-1}에 기반하여 3DoF 오디오 신호(들) x_{3DA}로부터 오디오 신호(들) x^*를 근사/복원한다
S906
근사된/복원된 오디오 신호 x^* 및 청취자 위치(및 선택적으로 VR 환경)에 기반하여 6DoF 오디오를 렌더링한다
S907