(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년04월05일
(11) 등록번호 10-2517906
(24) 등록일자 2023년03월30일

(51) 국제특허분류(Int. Cl.)
H04S 3/00 (2006.01) *A63F 13/211* (2014.01)
A63F 13/212 (2014.01) *A63F 13/213* (2014.01)
A63F 13/355 (2014.01) *A63F 13/428* (2014.01)
A63F 13/5255 (2014.01) *A63F 13/54* (2014.01)
G06F 3/01 (2006.01) *H04S 7/00* (2006.01)
(52) CPC특허분류
H04S 3/008 (2020.05)
A63F 13/211 (2015.01)
(21) 출원번호 10-2019-7023648
(22) 출원일자(국제) 2018년06월15일
심사청구일자 2021년06월09일
(85) 번역문제출일자 2019년08월12일
(65) 공개번호 10-2020-0018773
(43) 공개일자 2020년02월20일
(86) 국제출원번호 PCT/US2018/037887
(87) 국제공개번호 WO 2018/232327
국제공개일자 2018년12월20일
(30) 우선권주장
17176248.7 2017년06월15일
유럽특허청(EPO)(EP)
(뒷면에 계속)
(56) 선행기술조사문헌
US20170018121 A1
JP2015233252 A

(73) 특허권자
돌비 인터네셔널 에이비
아일랜드 디02 브이케이60 더블린 그랜드 커널 독 랜즈 블록 씨 서 존 로저슨스 키 77
돌비 레버러토리즈 라이센싱 코오포레이션
미합중국, 캘리포니아 94103, 샌프란시스코, 마켓 스트리트 1275
(72) 발명자
페르쉬, 크리스토프
미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레보레토리어스, 인크.내
칭고스, 니콜라스 알.
미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레보레토리어스, 인크.내
(74) 대리인
특허법인 광장리앤고

전체 청구항 수 : 총 21 항 심사관 : 권영학

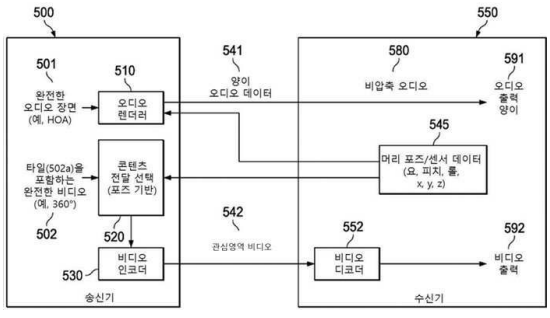
(54) 발명의 명칭 **컴퓨터 매개 현실 애플리케이션에서 송신기와 수신기 사이의 통신을 최적화하는 방법, 장치 및 시스템**

(57) 요 약

본 발명은 제1 장치에 의한 재현을 위하여 미디어 콘텐츠를 처리하는 시스템, 방법 및 장치에 관한 것이다. 방법은 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하는 것을 포함한다. 포즈 정보는 미디어 콘텐츠를 제공하는 제2 장치로 전송된다. 미디어 콘텐츠는 포즈 정보에 기반하여 렌더링되어 렌더링된 미디어 콘텐츠를 획

(뒷면에 계속)

대 표 도 - 도5



득한다. 렌더링된 미디어 콘텐츠는 재현을 위하여 제1 장치로 전송된다. 본 발명은 미디어 콘텐츠를 재현하는 제1 장치 및 미디어 콘텐츠를 저장하는 제2 장치를 포함할 수 있다. 제1 장치는 나타내는 포즈 정보를 획득하고 포즈 정보를 제2 장치로 전송하도록 구성되며; 제2 장치는 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링 하고; 재현을 위하여 렌더링된 미디어 콘텐츠를 제1 장치로 전송하도록 적응된다.

(52) CPC특허분류
A63F 13/212 (2015.01)
A63F 13/213 (2015.01)
A63F 13/355 (2015.01)
A63F 13/428 (2015.01)
A63F 13/5255 (2015.01)
A63F 13/54 (2015.01)
G06F 3/011 (2022.02)
G06F 3/012 (2013.01)
H04S 7/302 (2013.01)

(30) 우선권주장
62/519,952 2017년06월15일 미국(US)
62/680,678 2018년06월05일 미국(US)

명 세 서

청구범위

청구항 1

오디오 콘텐츠를 처리하는 방법에 있어서, 상기 방법은:

하나 이상의 프로세서를 사용하여, 사용자의 제1 위치 또는 방향을 나타내는 제1 포즈(pose) 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 제1 포즈 정보 및 이전의 포즈 정보에 기반하여 예측 포즈 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 렌더링된 오디오 콘텐츠를 획득하기 위해 상기 예측 포즈 정보에 기반하여 상기 오디오 콘텐츠를 렌더링하는 것; 및

상기 하나 이상의 프로세서를 사용하여, 상기 렌더링된 오디오 콘텐츠 및 예측 포즈 정보를 재현을 위한 제1 장치로 전송하는 것 - 상기 제1 장치에서 업데이트된 렌더링된 오디오 콘텐츠를 재현하기 전에, 상기 렌더링된 오디오 콘텐츠를 업데이트하도록 상기 예측 포즈 정보 및 제2 포즈 정보가 사용됨 - 을 포함하는 방법.

청구항 2

제1항에 있어서, 상기 오디오 콘텐츠는 일차 앰비소닉(First Order Ambisonics: FOA) 기반, 고차 앰비소닉(Higher Order Ambisonics: HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 둘 이상의 조합인 방법.

청구항 3

제1항에 있어서, 상기 렌더링된 오디오 콘텐츠는 양이(binaural) 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 둘 이상의 조합인 방법.

청구항 4

제1항에 있어서, 상기 렌더링된 오디오 콘텐츠를 업데이트하도록 상기 예측 포즈 정보 및 상기 제2 포즈 정보를 사용하는 것은,

상기 예측 포즈 정보를 상기 제2 포즈 정보와 비교하는 것; 및

상기 비교의 결과에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하는 것을 더 포함하는 방법.

청구항 5

제4항에 있어서, 상기 예측 포즈 정보는 상기 렌더링된 오디오 콘텐츠가 재현을 위한 상기 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측되며;

상기 제2 포즈 정보는 상기 렌더링된 오디오 콘텐츠가 재현을 위한 상기 제1 장치에 의해 실제로 처리되는 타이밍에 획득된 포즈 정보인 방법.

청구항 6

제1항에 있어서, 상기 렌더링된 오디오 콘텐츠는 비압축 형태로 상기 제1 장치로 전송되는 방법.

청구항 7

제1항에 있어서,

상기 제1 장치로의 전송 전에 상기 렌더링된 오디오 콘텐츠를 인코딩하는 것; 및 상기 제1 장치에서의 수신 후에 상기 인코딩된 렌더링된 오디오 콘텐츠를 디코딩하는 것을 더 포함하는 방법.

청구항 8

제5항에 있어서, 상기 렌더링된 오디오 콘텐츠가 재현을 위한 상기 제1 장치에 의해 처리될 것으로 예상되는 상기 타이밍의 추정은 상기 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 상기 렌더링된 오디오 콘텐츠를 상기 제1 장치로 전송하는 데 필요한 시간의 추정을 포함하는 방법.

청구항 9

제1항에 있어서, 상기 예측 포즈 정보는 상기 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 상기 렌더링된 오디오 콘텐츠를 상기 제1 장치로 전송하는 데 필요한 시간의 추정에 추가로 기반하여 획득되는 방법.

청구항 10

제1항에 있어서,

상기 오디오 콘텐츠를 렌더링하는 데 사용된 상기 예측 포즈 정보를 상기 제2 포즈 정보와 비교하는 것; 및

상기 비교의 결과에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하는 것을 더 포함하는 방법.

청구항 11

제1항에 있어서,

제2 장치에서, 상기 포즈 정보의 변화에 응답하여 상기 렌더링된 오디오 콘텐츠가 어떻게 변화하는지를 나타내는 기울기(gradient) 정보를 결정하는 것;

상기 렌더링된 오디오 콘텐츠와 함께 상기 기울기 정보를 상기 제1 장치로 전송하는 것;

상기 제1 장치에서, 상기 오디오 콘텐츠를 렌더링하는 데 사용된 상기 포즈 정보를 현재 포즈 정보와 비교하는 것; 및

상기 기울기 정보 및 상기 비교의 결과에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하는 것을 더 포함하는 방법.

청구항 12

제1항에 있어서,

상기 방법은 상기 제1 장치가 위치한 환경의 음향 특성(acoustic characteristics)을 나타내는 환경 정보를 제2 장치로 전송하는 것을 더 포함하고;

상기 오디오 콘텐츠의 렌더링은 상기 환경 정보에 추가로 기반하는 방법.

청구항 13

제1항에 있어서,

상기 방법은 상기 사용자 또는 상기 사용자의 일부의 형태를 나타내는 형태 정보(morphologic information)를 제2 장치로 전송하는 것을 더 포함하고;

상기 오디오 콘텐츠의 렌더링은 상기 형태 정보에 추가로 기반하는 방법.

청구항 14

시스템으로서:

하나 이상의 프로세서;

하나 이상의 프로세서에 의해 실행될 때, 상기 하나 이상의 프로세서로 하여금:

- 사용자의 제1 위치 또는 방향을 나타내는 제1 포즈 정보를 획득하는 것;

- 상기 제1 포즈 정보 및 이전의 포즈 정보에 기반하여 예측 포즈 정보를 획득하는 것;

- 렌더링된 오디오 콘텐츠를 획득하기 위해 상기 예측 포즈 정보에 기반하여 상기 오디오 콘텐츠를 렌더링하는 것; 및

- 상기 렌더링된 오디오 콘텐츠 및 상기 예측 포즈 정보를 재현을 위한 제1 장치로 전송하는 것 - 상기 제1 장치에서 업데이트된 렌더링된 오디오 콘텐츠를 재현하기 전에, 상기 렌더링된 오디오 콘텐츠를 업데이트하도록 상기 예측 포즈 정보 및 제2 포즈 정보가 사용됨 - 을 포함하는 동작을 수행하도록 하는 명령어를 저장하는 메모리를 포함하는 시스템.

청구항 15

제14항에 있어서, 상기 렌더링된 오디오 콘텐츠는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 둘 이상의 조합인 시스템.

청구항 16

제1 장치에 의한 재현을 위해 미디어 콘텐츠를 처리하는 방법으로서, 상기 방법은:

하나 이상의 프로세서를 사용하여, 사용자의 위치 또는 방향을 나타내는 포즈 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 렌더링된 미디어 콘텐츠를 획득하기 위해 상기 예측 포즈 정보에 기반하여 상기 미디어 콘텐츠를 렌더링하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 렌더링된 미디어 콘텐츠가 상기 포즈 정보의 변화에 응답하여 어떻게 변화하는지를 나타내는 기울기 정보를 획득하는 것; 및

상기 하나 이상의 프로세서를 사용하여, 상기 제1 장치에 의해 업데이트된 렌더링된 미디어 콘텐츠를 업데이트하고 재현하는 데 사용하기 위해 상기 렌더링된 미디어 콘텐츠 및 기울기 정보를 상기 제1 장치로 전송하는 것

을 포함하는, 방법.

청구항 17

제16항에 있어서,

상기 하나 이상의 프로세서를 사용하여, 상기 제1 장치에 의해 상기 업데이트된 렌더링된 미디어 콘텐츠를 업데이트하고 재현하는 데 사용하기 위해 상기 예측 포즈를 상기 제1 장치로 전송하는 것을 더 포함하는, 방법.

청구항 18

제1 장치의 하나 이상의 프로세서를 사용하여, 사용자의 제1 위치 또는 방향을 나타내는 제1 포즈 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 제1 포즈 정보를 제2 장치로 전송하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 제2 장치로부터 렌더링된 오디오 콘텐츠를 수신하는 것 - 상기 렌더링된 오디오 콘텐츠는 예측 포즈 정보를 사용하여 상기 제2 장치에 의해 렌더링되었으며, 상기 예측 포즈 정보는 상기 제1 포즈 정보에 기반함 - ;

상기 하나 이상의 프로세서를 사용하여, 상기 사용자의 제2 위치 또는 방향을 나타내는 제2 포즈 정보를 획득하는 것;

상기 하나 이상의 프로세서를 사용하여, 상기 제2 포즈 정보에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하는 것; 및

상기 하나 이상의 프로세서를 사용하여, 상기 업데이트된 렌더링된 오디오 콘텐츠를 재현하는 것을 포함하는 방법.

청구항 19

제18항에 있어서,

상기 하나 이상의 프로세서를 사용하여, 상기 제2 장치로부터 상기 예측 포즈를 수신하는 것; 및

상기 하나 이상의 프로세서를 사용하여, 상기 예측 포즈와 상기 제2 포즈 정보의 차이에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하는 것

을 더 포함하는, 방법.

청구항 20

렌더링된 오디오 콘텐츠를 재현하기 위한 제1 장치 및 상기 렌더링된 오디오 콘텐츠를 만들기 위한 제2 장치를 포함하는 시스템으로서,

상기 제1 장치는:

사용자의 제1 위치 또는 방향을 나타내는 제1 포즈 정보를 획득하고;

상기 제1 포즈 정보를 상기 제2 장치로 전송하고;

상기 제2 장치로부터 렌더링된 오디오 콘텐츠를 수신하고;

상기 사용자의 제2 위치 또는 방향을 나타내는 제2 포즈 정보를 획득하고;

적어도 상기 제2 포즈 정보를 사용하여 상기 렌더링된 오디오 콘텐츠를 업데이트하고;

상기 업데이트된 렌더링된 오디오 콘텐츠를 재현하도록 구성되는, 시스템.

청구항 21

제20항에 있어서, 상기 제1 장치는 상기 제1 장치로부터 예측 포즈를 수신하고 상기 예측 포즈와 상기 제2 포즈 정보의 차이에 기반하여 상기 렌더링된 오디오 콘텐츠를 업데이트하도록 더 구성되는, 시스템.

청구항 22

삭제

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

발명의 설명

기 술 분 야

[0001] 연관된 출원에 대한 상호참조

[0002] 이 출원은 2018년 6월 5일에 출원된 미국 특허 가출원 제62/680,678호; 2017년 6월 17일에 출원된 미국 특허 가출원 제62/519,952호 및 유럽 특허 출원 제17176248.7호의 우선권을 주장하며, 이들은 모두 전체로서 참조로 포함된다.

[0003] 기술분야

[0004] 본 개시는, 예를 들어, 가상 현실(VR), 증강 현실(AR) 및 혼합 현실(MR) 애플리케이션과 같은 컴퓨터 매개 현실 애플리케이션에 관한 것이다. 이들 애플리케이션은 클라이언트/수신기 양이화(binauralized) 및 비양이화(non-binauralized) 오디오 및 비디오 애플리케이션을 포함하지만 이에 제한되지 않는다.

배 경 기 술

[0005] 컴퓨터 매개 현실 공간(예를 들어, VR, AR 및 MR 공간)의 애플리케이션 및 제품은 사운드 소스 및 장면(scene)에 대한 점점 더 세련된 음향 모델을 포함하도록 급속하게 진화하고 있다. 제한을 의도하지 않고, 이 문서의 나머지 부분에서 VR, AR 및 MR을 참조한다. 컴퓨터 매개 현실 경험을 최적화하기 위하여, 사용자 움직임(예를 들어, 머리 움직임)과 이 움직임에 적응된 사운드(렌더링된 사운드) 사이의 지연을 최소화하는 것이 바람직하다. 이러한 지연은 모션 투 사운드 대기시간(motion-to-sound latency) 또는 모션 투 이어 지체(motion-to-ear lag)로도 알려져 있다. 또한, 계산 복잡도 및 전력 소모를 최적화하는 것이 중요한, 스마트폰과 같은 공통 수신기 디바이스에 대해 사운드를 디코딩하고 렌더링하는 데 필요한 명령의 수를 최소화하는 것이 또한 바람직하다. 예를 들어, 비통신(non-communication) 사례에 대해 전체 오디오 장면이 전송되면, 수신기의 렌더링 대기시간에 초점이 맞추어진다. 예를 들어 선형 애플리케이션(예를 들어, 영화)은 사용자의 동작에 동적으로 반응하지 않는다. 그러나, 대화형 콘텐츠의 경우 모든 누적 왕복 대기시간이 고려되어야 한다(예를 들어, 사용자가 렌더링을 위해 서버로 다시 보내야 하는 이벤트를 트리거링하는 경우). 동적으로 변화되는 콘텐츠는, 콘텐츠가 소비되기 전에, 사용자가 모션과 그 결과적인 효과 사이의 대기시간을 인식하지 못하고 콘텐츠의 오디오와 비디오 사이에 오정렬이 없도록 충분한 리드(lead) 타임으로 인코딩되어야 한다. 사용자 움직임(위치 및/또는 방향)이 콘텐츠 자체에 영향을 미치지 않으므로, 선형 애플리케이션의 경우 모션 투 사운드 대기시간에 대해 인코딩 및 디코딩 대기시간을 고려하지 않는다. 오히려, 이러한 움직임은 콘텐츠를 보는 관점에만 영향을 미친다. 따라서, 선형 콘텐츠의 경우, 사용자 움직임은 렌더링에만 영향을 미칠 뿐, 출력 사운드의 인코딩 및/또는 디코딩에는 영향을 미치지 않는다. 통신 사례는 콘텐츠(예를 들어, 대사)가 발생할 때에만 시스템이 매체 인코딩, 전송 및 디코딩을 시작할 수 있기 때문에 차이가 있다. (예를 들어, 게임 엔진으로부터) 대화형 콘텐츠가 실시간으로 렌더링되고 클라우드의 원격 서버에 의해 인코딩되는 경우에도 마찬가지이다. 또한, 이 차이는 멀미(motion-sickness)를 일으킬 수 있어, 비디오 및 오디오 시스템의 전체 대기시간이 동일한 것이 매우 중요하다. 따라서 비디오 시스템의 대기시간에 따라, 유사한 레벨의 오디오 시스템 대기시간을 달성할 필요가 있다.

발명의 내용

[0006] 본 문서는 강력한(compelling) 경험을 전달하기 위해 너무 높은 대기시간과 너무 높은 계산 복잡도를 요구하는 일반적인 AR, VR 및 MR 시스템의 기술적인 문제점을 해결하고자 한다. 이 문제를 해결하기 위해, 본 문서는 각 독립항의 특징을 갖는, 미디어 콘텐츠를 처리하는 방법, 미디어 콘텐츠를 처리하는 시스템 및 대응하는 장치를

제안한다.

[0007] 본 개시의 일 양상은 제1 장치에 의한 재현(reproduction)을 위하여 미디어 콘텐츠를 처리하는 방법에 관한 것이다. 제1 장치는 예를 들어 수신기, 수신기 장치 또는 재생 장치 중 하나일 수 있다. 상기 제1 장치는 예를 들어 AR/VR/MR 헤드셋과 같은 AR/VR/MR 장비에 대응하거나, 이를 포함하거나 또는 이와 함께 작동할 수 있다. 이에 따라, 제1 장치는 미디어 콘텐츠를 재현하기 위한 재현 장비(예를 들어, 스피커, 헤드폰) 및 재현 장비에 연결되는 프로세서를 포함할 수 있다. 미디어 콘텐츠는 오디오 콘텐츠 및/또는 비디오 콘텐츠이거나 이를 포함할 수 있다. 처리는 렌더링을 수반하거나 이에 해당할 수 있다. 재현은 재생(replay)을 수반하거나 이에 해당할 수 있다. 방법은 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하는 것을 포함할 수 있다. 포즈 정보를 획득하는 것은 제1 장치에서 수행될 수 있다. 사용자는 제1 장치의 사용자일 수 있다. 포즈 정보는 예를 들어 사용자의 머리에 연관될 수 있다. 포즈 정보는 사용자와 등록되어 배열될 수 있는 센서, 예를 들어, 포즈 센서에 의해 획득될 수 있다. 따라서, 포즈 정보는 센서 데이터로 지칭될 수 있다. 포즈 정보는 포즈의 하나 이상의 일차 미분 및/또는 포즈의 하나 이상의 이차 미분을 더 포함할 수 있다. 사용자는 예를 들어 AR/VR/MR 장비를 착용할 수 있다. 방법은 미디어 콘텐츠를 제공(예를 들어, 저장, 재생)하는 제2 장치로 포즈 정보를 전송하는 것을 더 포함할 수 있다. 제2 장치는 예를 들어 송신기 장치, 서버 장치 또는 콘텐츠 전달 장치 중 하나일 수 있다. 제2 장치는 미디어 콘텐츠를 제1 장치로 제공하는 장치일 수 있다. 제1 및 제2 장치는 서로 공간적으로 분리될 수 있다. 방법은 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하는 것을 더 포함할 수 있다. 렌더링은 제2 장치에서 수행될 수 있다. 렌더링된 미디어 콘텐츠는 사전 렌더링된 미디어 콘텐츠로 지칭될 수 있다. 렌더링은 예를 들어 오디오 콘텐츠의 경우 둘 이상의 채널일 수 있다. 방법은 렌더링된 미디어 콘텐츠를 재현을 위하여 제1 장치로 전송하는 것을 더 포함할 수 있다. 렌더링된 미디어 콘텐츠를 전송하는 것은 제2 장치에 의해 수행될 수 있다. 방법은 (제1 장치에 의해) 렌더링된 미디어 콘텐츠를 재현(예를 들어, 재생)하는 것을 추가로 더 포함할 수 있다.

[0008] 렌더링된 미디어 콘텐츠만이 송신될 필요가 있는 경우, 무손실 데이터 전송을 위한 전송 비트율은 완전한 미디어 콘텐츠의 압축 버전의 비트율과 비슷하거나 유사할 수 있다. 따라서, 제안된 방법의 문맥에서 압축은 필수적이지 않을 수 있다. 비압축 또는 무손실 미디어 스트림을 전송하는 것은 인코딩 및 디코딩을 위한 대기시간을 제거하거나 줄일 것이다. 예를 들어, 인코딩/디코딩으로 인한 대기시간은 0으로 감소될 수 있고, 이는 모션 투 이어 대기시간 및/또는 모션 투 아이(motion-to-eye) 대기시간의 전반적인 감소를 가져온다. 또한 사전 렌더링된 미디어 콘텐츠의 압축이 없을 때 제1 장치(수신기)는 디코딩 또는 렌더링 없이 오디오/비디오를 출력할 수 있다. 이 경우 디코딩을 수행할 필요가 없으며 및/또는 송신기 측에서 렌더링이 이미 완료되었으므로 수신기에서의 계산 복잡도 감소를 가져올 것이다. 따라서, 제안된 방법은 모션 투 이어 대기시간 및/또는 모션 투 아이 대기시간을 줄일 수 있고, 또한 수신기 측에서의 계산 복잡도를 줄일 수 있다.

[0009] 일부 실시예에서, 미디어 콘텐츠는 오디오 콘텐츠를 포함할 수 있고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함할 수 있다. 대안적으로, 또는 부가적으로, 미디어 콘텐츠는 비디오 콘텐츠를 포함할 수 있고 렌더링된 미디어 콘텐츠는 렌더링된 비디오 콘텐츠를 포함할 수 있다.

[0010] 일부 실시예에서, 미디어 콘텐츠는 오디오 콘텐츠를 포함할 수 있고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함할 수 있다. 그러면, 방법은 렌더링된 오디오 콘텐츠의 가청(예를 들어, 음향) 표현을 생성하는 것을 더 포함할 수 있다. 가청 표현(audible representation)의 생성은 제1 장치에서 수행될 수 있다. 상기 생성은 예를 들어 오디오 콘텐츠의 경우 제1 장치의 둘 이상의 라우드스피커(loudspeaker)를 통해 수행될 수 있다.

[0011] 일부 실시예에서, 오디오 콘텐츠는 일차 앰비소닉(First Order Ambisonics: FOA) 기반, 고차 앰비소닉(Higher Order Ambisonics: HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합일 수 있다.

[0012] 일부 실시예에서, 렌더링된 오디오 콘텐츠는 양이(binaural) 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합일 수 있다.

[0013] 일부 실시예에서, 렌더링은 포즈 정보에 기반하고 이전 포즈 정보 및/또는 하나 이상의 일차 및/또는 이차 미분에 추가로 기반하여 예측 포즈 정보를 획득하는 것을 수반할 수 있다. 예측 포즈 정보는 미래 타이밍에 대한 포즈 정보일 수 있다. 이전 포즈 정보는 이전 타이밍에 제1 장치에서 획득되거나 이로부터 수신된 포즈 정보일 수 있다. 예측은 제2 장치에서 수행될 수 있다. 대안적으로, 예측이 제1 장치에서 수행될 수 있다. 후자의 경우,

제1 장치가 예측 포즈 정보를 제2 장치로 전송할 수 있다. 렌더링은 렌더링된 미디어 콘텐츠를 획득하기 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하는 것을 더 수반할 수 있다.

[0014] 예측 포즈 정보를 고려함으로써, 렌더링된 미디어 콘텐츠의 인코딩/디코딩 및/또는 렌더링된 미디어 콘텐츠의 제1 장치로의 전송으로 인한 지연이 고려될 수 있다. 달리 말하자면, 상기 지연이 예측 포즈 정보에 대해 적절히 감추어질 수 있어 사용자는 이 지연을 알지 못하게 되거나 오디오, 비디오 및 움직임 사이의 불일치를 인지하지 않을 수 있다.

[0015] 일부 실시예에서, 방법은 렌더링된 미디어 콘텐츠와 함께 예측 포즈 정보를 제1 장치로 전송하는 것을 더 포함할 수 있다.

[0016] 이는 제1 장치가 예측 포즈 정보(즉, 이 경우 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보)가 실제/현재 포즈 정보(즉, 제1 장치에서 현재 획득된 포즈 정보)와 동일(또는 실질적으로 동일)한지 확인을 수행하고, 예측 포즈 정보와 실제/현재 포즈 정보 사이에 불일치가 있으면 렌더링된 미디어 콘텐츠를 적절히 적응할 수 있게 한다.

[0017] 일부 실시예에서, 방법은 예측 포즈 정보를 실제 포즈 정보와 비교하는 것을 더 포함할 수 있다. 방법은 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 추가로 더 포함할 수 있다. 상기 비교 및 상기 업데이트는 제1 장치에서 수행될 수 있다. 실제 포즈 정보는 예를 들어 렌더링된 미디어 콘텐츠가 제1 장치에 의해 재현되는 타이밍에서의(예를 들어, 타이밍에 획득되는) 포즈 정보일 수 있다. 업데이트는 예를 들어 예측 포즈 정보와 실제 포즈 정보 사이의 차이에 기반하여 수행될 수 있다. 상기 업데이트는 렌더링된 미디어 콘텐츠의, 예를 들어 회전, 레벨 변경 및/또는 블라인드 업믹싱(blind upmixing)에 의한 외삽을 수반할 수 있다.

[0018] 일부 실시예에서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측될 수 있다. 제1 장치에 의한 렌더링된 미디어 콘텐츠의 처리는 렌더링된 미디어 콘텐츠의 재현(예를 들어, 재생)을 수반할 수 있다. 실제 포즈 정보(예를 들어, 현재 포즈 정보)는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 실제로 처리되는 타이밍에 획득되는 포즈 정보일 수 있다. 실제 포즈 정보는 렌더링된 미디어 콘텐츠가 제1 장치에 의해 실제로 처리되는 타이밍에 획득될 수 있다.

[0019] 따라서, 예측 포즈 정보와 실제 포즈 정보 사이의 임의의 불일치가 고려될 수 있고 이에 따라 렌더링된 미디어 콘텐츠를 사용자의 포즈(예를 들어, 사용자 머리의 포즈)에 더 잘 적응시킬 수 있으며 사용자에 대해 인지되고 예상되는 오디오/비디오 장면의 임의의 괴리를 피할 수 있다. 예측 포즈 정보와 실제 포즈 정보 사이의 불일치가 작을 것으로 예상되므로, 이러한 적응은 감당할 만한 계산 복잡도로 안전하게 제1 장치에 맡겨질 수 있다.

[0020] 일부 실시예에서, 렌더링된 미디어 콘텐츠는 비압축 형태로 제1 장치로 전송될 수 있다.

[0021] 이는 제1 장치(수신기)에서의 계산 복잡도를 줄이고, 또한 포즈의 변화와 변화된 포즈에 따라 렌더링된 미디어 콘텐츠의 재현 사이의 왕복 지연을 줄인다.

[0022] 일부 실시예에서, 방법은 제1 장치로의 전송 전에 렌더링된 미디어 콘텐츠를 인코딩(예를 들어, 압축)하는 것을 더 포함할 수 있다. 방법은 또한 제1 장치에서의 수신 후에 인코딩된 렌더링된 미디어 콘텐츠를 디코딩(예를 들어, 압축해제)하는 것을 더 포함할 수 있다. 인코딩/디코딩은 렌더링된 미디어 콘텐츠의 압축/압축해제를 수반하거나 이에 대응할 수 있다. 인코딩/디코딩은 저지연(LowDelay) 인코딩/디코딩일 수 있다.

[0023] 일부 실시예에서, 렌더링된 오디오 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정은 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정을 포함할 수 있다.

[0024] 일부 실시예에서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정에 추가로 기반하여 획득될 수 있다.

[0025] 따라서, 인코딩/디코딩 및/또는 전송으로 인한 지연이 장면 내에 감추어질 수 있고 사용자가 이들 지연을 알지 못하게 된다.

[0026] 일부 실시예에서, 방법은 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하는 것을 더 포함할 수 있다. 현재 포즈 정보는 예를 들어 렌더링된 미디어 콘텐츠를 재현하는 시간에 획득된 포즈 정보일 수 있다. 방법은 또한 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 더 포함할 수 있다. 업데이트는 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보와 현재 포즈 정보 사이의 차이에 기반하여

수행될 수 있다. 상기 업데이트는 렌더링된 미디어 콘텐츠의, 예를 들어 회전, 레벨 변경 및/또는 블라인드 업믹싱에 의한 외삽(extrapolation)을 수반할 수 있다.

[0027] 일부 실시예에서, 방법은, 제2 장치에서, 포즈 정보의 변화(예를 들어, 포즈의 변화)에 응답하여 렌더링된 미디어 콘텐츠가 어떻게 변화하는지를 나타내는 기울기(gradient) 정보를 결정하는 것을 더 포함할 수 있다. 기울기 정보는 (오디오 콘텐츠에 대하여) 사용자(예를 들어, 사용자의 머리)의 병진 및/또는 회전에 응답하는 (예를 들어, 각 채널의) 부대역(sub-band) 에너지 레벨의 변화를 나타낼 수 있다. 방법은 렌더링된 미디어 콘텐츠와 함께 기울기 정보를 제1 장치로 전송하는 것을 더 포함할 수 있다. 방법은, 제1 장치에서, 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하는 것을 더 포함할 수 있다. 미디어 콘텐츠를 렌더링하는 데 (제2 장치에 의해) 사용된 포즈 정보는 렌더링된 미디어 콘텐츠와 함께 제1 장치로 전송될 수 있다. 이 포즈 정보가 렌더링된 미디어 콘텐츠와 함께 제1 장치로 전송되지 않는 경우, 제1 장치는 제2 장치로 송신한 포즈 정보를 참조할 수 있다. 현재 포즈 정보는 예를 들어 렌더링된 미디어 콘텐츠를 재현하는 시간에 획득된 포즈 정보일 수 있다. 방법은 또한 기울기 정보 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 더 포함할 수 있다. 렌더링된 미디어 콘텐츠의 업데이트는 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보와 현재 포즈 정보 사이의 차이에 기반하여 수행될 수 있다. 상기 업데이트는 렌더링된 미디어 콘텐츠의, 예를 들어 회전, 레벨 변경 및/또는 블라인드 업믹싱에 의한 외삽을 수반할 수 있다.

[0028] 이에 의해, 포즈 정보 예측에서의 작은 결함이 정정될 수 있으며 포즈와 재현된 미디어 콘텐츠 사이의 임의의 불일치를 피할 수 있다.

[0029] 일부 실시예에서, 미디어 콘텐츠는 오디오 콘텐츠를 포함할 수 있고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함할 수 있다. 그러면, 방법은 제1 장치가 위치한 환경의 음향 특성(acoustic characteristics)을 나타내는 환경 정보를 제2 장치로 전송하는 것을 더 포함할 수 있다. 이 경우, 미디어 콘텐츠의 렌더링은 환경 정보에 추가로 기반할 수 있다. 환경 정보는 룸 특성 및/또는 양이 룸 임펄스 응답(Binaural Room Impulse Response: BRIR) 함수를 포함할 수 있다.

[0030] 이는 재현된 미디어 콘텐츠를 사용자가 위치한 특정 환경에 구체적으로 적응할 수 있게 하여, 사용자의 컴퓨터 매개 현실 경험을 향상시킨다.

[0031] 일부 실시예에서, 미디어 콘텐츠는 오디오 콘텐츠를 포함할 수 있고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함할 수 있다. 그러면, 방법은 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보(morphologic information)를 제2 장치로 전송하는 것을 더 포함할 수 있다. 이 경우, 미디어 콘텐츠의 렌더링은 형태 정보에 추가로 기반할 수 있다. 형태는 형상 또는 크기, 예를 들어, 사용자 머리의 형상 또는 크기를 포함하거나 이에 대응할 수 있다. 형태 정보는 머리 전달 함수(Head-Related Transfer Function: HRTF)를 포함할 수 있다. 렌더링은 양이 렌더링일 수 있다.

[0032] 이는 재현된 미디어 콘텐츠를 사용자 또는 사용자의 일부의 특정 형태에 구체적으로 적응할 수 있게 하여, 사용자의 컴퓨터 매개 현실 경험을 향상시킨다.

[0033] 본 개시의 다른 양상은 위의 양상 및 그 실시예에 따른(예를 들어, 구현하는) 제1 장치, 제2 장치, 및 제1 장치와 제2 장치의 시스템에 관한 것이다.

[0034] 따라서, 본 개시의 다른 양상은 미디어 콘텐츠를 재현하는 제1 장치 및 미디어 콘텐츠를 저장하는 제2 장치를 포함하는 시스템에 관한 것이다. 제1 장치는 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하도록 적응(구성)될 수 있다. 제1 장치는 포즈 정보를 제2 장치로 전송하도록 더 적응(구성)될 수 있다. 제2 장치는 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하도록 적응(구성)될 수 있다. 제2 장치는 렌더링된 미디어 콘텐츠를 재현하기 위하여 제1 장치로 전송하도록 더 적응(구성)될 수 있다. 예를 들어, 제1 및 제2 장치는 각각의 프로세서(또는 각각의 프로세서 세트) 및 각각의 프로세서(또는 각각의 프로세서 세트)에 연결되는 메모리를 포함할 수 있다. 프로세서는 위에서 설명된 작업을 수행하도록 적응(구성)될 수 있다.

[0035] 본 개시의 다른 양상은 제1 장치에 의한 재현을 위하여 미디어 콘텐츠를 제공하는 제2 장치에 관한 것이다. 제2 장치는 제1 장치의 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 수신하도록 적응(구성)될 수 있다. 제2 장치는 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하도록 더 적응(구성)될 수 있다. 제2 장치는 또한 렌더링된 미디어 콘텐츠를 재현을 위하여 제1 장치로 전송하도록 더 적응(구성)될 수 있다. 예를 들어, 제2 장치는 프로세서(또는 프로세서 세트) 및 프로세서(또는 프로세서 세트)에

연결되는 메모리를 포함할 수 있다. 프로세서(또는 프로세서 세트)는 위에서 설명된 작업을 수행하도록 적응(구성)될 수 있다.

[0036] 본 개시의 다른 양상은 제2 장치에 의해 제공되는 미디어 콘텐츠를 재현하는 제1 장치에 관한 것이다. 제1 장치는 제1 장치의 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하도록 적응(구성)될 수 있다. 제1 장치는 포즈 정보를 제2 장치로 전송하도록 더 적응(구성)될 수 있다. 제1 장치는 제2 장치로부터 렌더링된 미디어 콘텐츠를 수신하도록 더 적응(구성)될 수 있다. 렌더링된 미디어 콘텐츠는 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하여 획득될 수 있다. 제1 장치는 또한 렌더링된 미디어 콘텐츠를 재현하도록 더 적응(구성)될 수 있다. 예를 들어, 제1 장치는 프로세서(또는 프로세서 세트) 및 프로세서(또는 프로세서 세트)에 연결되는 메모리를 포함할 수 있다. 프로세서(또는 프로세서 세트)는 위에서 설명된 작업을 수행하도록 적응(구성)될 수 있다.

[0037] 방법에 대해 이루어진 임의의 설명은 이러한 방법/시스템에서 사용되는 대응하는 시스템 및 장치에 마찬가지로 적용되며, 그 반대도 마찬가지이다.

[0038] 본 개시의 또 다른 양상은, 송신기(S) 장치에 의하여, 사용자 위치 및/또는 방향 데이터를 수신하는 것 및, 통상적으로 객체 기반 또는 FOA/HOA 표현으로부터 유도되는, 대응하는 사전 렌더링된 콘텐츠를 송신하는 것을 포함하는 오디오 콘텐츠를 렌더링하는 방법을 수행하도록 구성되는 시스템, 장치, 방법 및 컴퓨터 판독 가능 저장 매체에 관한 것이다. 송신기에 의해 생성되는 사전 렌더링된 신호는 양이, FOA, HOA 또는 임의 유형의 채널 기반 렌더링이다. 방법은 비압축 사전 렌더링된 콘텐츠를 전송하는 것을 더 포함할 수 있다. 방법은 사전 렌더링된 콘텐츠를 인코딩하는 것 및 인코딩된 사전 렌더링된 콘텐츠를 전송하는 것을 더 포함할 수 있다. 방법은 수신기에 의하여 사전 렌더링된 콘텐츠를 수신하는 것을 더 포함할 수 있다. 방법은 수신기에 의해 사전 렌더링된, 사전 인코딩된 양이화된 콘텐츠를 디코딩하는 것을 더 포함할 수 있다. 사용자 위치 및/또는 방향 데이터는 세계 공간 내의 사용자의 위치 및 방향을 나타내는 로컬 포즈를 포함할 수 있다. 사용자 위치 데이터는 수신기로부터 송신기로 전송될 수 있다. 방법은 사전 렌더링된 양이화된 콘텐츠를 위해 사용된 사용자 위치 데이터를 수신기로 다시 전송하는 것을 더 포함할 수 있다. 방법은 수신된 사용자 위치 데이터 및 로컬 위치 데이터에 기반하여 사전 렌더링된 콘텐츠를 외삽하여 업데이트된 콘텐츠를 결정하는 것을 더 포함할 수 있다. 방법은 개별화된 양이 처리를 위하여 사용자에 대한 형태 데이터(예를 들어 머리 크기)를 전송하는 것을 더 포함할 수 있다. 방법은 BRIR 및 룸 특성에 대한 데이터를 전송하는 것을 더 포함할 수 있다. 방법은 콘텐츠가 청취자를 알 수 없는 방식(예를 들어, HRTF를 포함하지 않음)으로 전송된다는 결정에 기반하여, 수신기 측에서 양이 렌더링 및 개별화를 수행하는 것을 더 포함할 수 있다. 방법은 시점 t1에서 사용자 위치 및/또는 방향 데이터 P(t0)를 제공하는 것을 더 포함할 수 있다. 비압축 사전 렌더링된 콘텐츠는 양이화된 비압축 사전 렌더링된 콘텐츠일 수 있다.

도면의 간단한 설명

[0039] 첨부된 도면을 참조하여 이하에서 본 개시의 실시예를 설명한다.

도 1은 수신기의 제1 예를 도시한다.

도 2는 수신기의 제2 예를 도시한다.

도 3은 수신기 및 서버 시스템의 제1 예를 도시한다.

도 4는 송신기 및 수신기 시스템의 제2 예를 도시한다.

도 5는 송신기 및 수신기 시스템의 제3 예를 도시한다.

도 6는 송신기 및 수신기 시스템의 제4 예를 도시한다.

도 7은 미디어 콘텐츠 처리 방법의 제1 예를 도시한다.

도 8은 미디어 콘텐츠 처리 방법의 제2 예를 도시한다.

도 9는 미디어 콘텐츠 처리 방법의 제3 예를 도시한다.

도 10은 미디어 콘텐츠 처리 방법의 제4 예를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0040] 당업자라면 이해할 수 있는 바와 같이, 가상 세계에서의 완전한 몰입은 사람의 두뇌가 감지되는 것을 믿도록 "속인다(trick)". 시야가 시야 범위에 의해 제한되면, 소리는 눈에 보이지 않는 것(예를 들어, 뒤에서 돌진하는 황소, 오른쪽의 방울뱀, 왼쪽 귀에서 오른쪽 귀를 향해 머리 뒤에서 움직이는 속삭임조차도)에 차원을 추가한다. 따라서, 콘텐츠 제작자는 소리를 활용하여 사용자의 시선을 유도함으로써 효과적으로 이야기를 전달할 수 있다. 몰입형 오디오 경험은 현재 객체 또는 일차/고차 앰비소닉(First/Higher Order Ambisonics: FOA/HOA) 기반의 사운드 생성, 패키징 및 콘텐츠 재생을 통해 시네마 및 홈시어터에서 제공되고 있다. VR 사운드는 가상 세계에서 완벽한 몰입을 위해 정확한 사운드가 있어야 한다. VR 콘텐츠 제작자는 3차원 공간에서 객체 및/또는 HOA 기반 사운드를 생성할 수 있는 기능을 요구한다. 또한, 이러한 콘텐츠는 사용자가 콘텐츠를 즐길 수 있는 정확성과 효율성으로 인코딩, 전달, 디코딩 및 양이적으로 (헤드폰 또는 라우드스피커를 통해) 렌더링되어야 한다.

[0041] 수신기는 다양한 매개변수, 예컨대 대역폭 및 미디어 비트율에 기반하여, MPEG-DASH 또는 MPEG-MMT 포맷을 통해 전달된 오버더톱(over-the-top: OTT) 콘텐츠와 같은 콘텐츠의 미디어 포맷 표현을 선택할 수 있다. 수신기는 또한 미디어 소비에 관한 정보를 수신할 수 있다. 미디어 포맷 표현의 선택은 이러한 미디어 소비에 기반할 수 있다. 예를 들어, 사전 렌더링된 양이화된 데이터는 (예를 들어, 크로스토크 제거를 갖춘) 헤드폰 또는 스테레오 라우드스피커 출력의 표시에 기반하여 선택될 수 있다.

[0042] 본원에 기술된 예시적인 실시예는 미디어 콘텐츠를 처리(예를 들어, 오디오 콘텐츠 렌더링)하도록 적응되는 방법, 장치 및 프로세스를 설명한다. 예시적인 실시예가 일반적으로 (예를 들어, 오디오 콘텐츠 및/또는 비디오 콘텐츠를 포함하는) 미디어 콘텐츠를 처리하는 것에 관한 것이지만, 제한을 의도하지 않고, 본 문서의 나머지 부분에서 오디오 콘텐츠에 대한 참조가 이루어질 수 있다.

[0043] 도 1은 양이(binaural) 렌더링을 위한 수신기/클라이언트 시스템(100)의 예를 도시한다. 시스템 (100)은 오디오 입력(101)을 수신할 수 있다. 오디오 입력(101)은 송신기로부터의 인코딩된 비트스트림에 포함된 전체 장면(scene)을 포함할 수 있다. 수신기 시스템(100)은 사용자 움직임 및/또는 사용자 머리 방향에 관한 센서 데이터(포즈 정보)(110)를 수신 또는 검출할 수 있다. 센서 데이터(110)는, 예를 들어, 요, 피치, 롤 및/또는 (x, y, z) 좌표와 같은 방향 및 위치에 관한 정보를 포함할 수 있다. 수신기 시스템(100)은 오디오 입력(101)을 비압축 오디오 및/또는 메타데이터(120)로 디코딩할 수 있는 디코더(102)를 더 포함할 수 있다. 수신기 시스템(100)은 비압축 오디오 및/또는 메타데이터(120)를 양이 출력(150)으로 렌더링할 수 있는 렌더러(103)를 더 포함할 수 있다. 수신기 시스템(100)은 예를 들어 헤드폰 출력으로 양이 출력(150)을 출력할 수 있다.

[0044] 도 1에 도시된 수신기/클라이언트 시스템(100)은 이 문서의 시작 부분에서 설명된 대기시간 및/또는 계산 복잡도와 관련된 문제로 어려움을 겪을 수 있다.

[0045] 이 문제를 해결하기 위하여, 본 개시는, (예를 들어, 오디오 및/또는 비디오 콘텐츠를 포함하는) 미디어 콘텐츠를 처리하는 시스템 내에서, 수신기에서 사용자에 대한 포즈 정보를 획득하고, 송신기로 포즈 정보를 전송하고, 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하고, 렌더링된 미디어 콘텐츠를 수신기로 전송하도록 제안한다. 이에 따라, 수신기 측에서 수행되어야 하는 연산의 계산 복잡도가 현저히 줄어들 수 있다. 또한, 렌더링된 미디어 콘텐츠는 비압축 형태로 전송될 수 있어, 포즈의 변화(예를 들어, 머리 움직임)와 이러한 포즈의 변화에 적응된 재현된 미디어 콘텐츠의 인지(예를 들어, 사운드의 인지) 사이의 지연을 줄일 수 있다.

[0046] 도 7은 위의 고려사항에 따라 미디어 콘텐츠를 처리하는 방법(700)의 예를 개략적으로 도시하는 흐름도이다. 미디어 콘텐츠는 오디오 콘텐츠 및/또는 비디오 콘텐츠를 포함할 수 있다. 오디오 콘텐츠는 예를 들어 FOA 기반 오디오 콘텐츠, HOA 기반 오디오 콘텐츠, 객체 기반 오디오 콘텐츠, 채널 기반 오디오 콘텐츠 또는 그 조합일 수 있다. 미디어 콘텐츠의 처리는 미디어 콘텐츠 렌더링을 수반할 수 있다. 방법은 미디어 콘텐츠를 재현하는 제1 장치 및 미디어 콘텐츠를 제공하는 제2 장치를 포함하는 시스템 내에서 수행될 수 있다. 미디어 콘텐츠의 재현은 미디어 콘텐츠의 재생을 수반할 수 있다. 제1 장치는 수신기, 수신기 장치, 클라이언트, 클라이언트 장치 또는 재생 장치로 지칭될 수 있으며, 제1 장치는, 예를 들어, VR/AR/MR 헤드셋(예를 들어, 고글)과 같은 컴퓨터 매개 현실(예를 들어, VR, AR, MR) 장비를 포함하거나, 이에 대응하거나 또는 이와 함께 작동할 수 있고, 사용자와 연관될 수 있다. 사용자는 컴퓨터 매개 현실 장비를 착용할 수 있다. 제1 장치는 사용자 또는 사용자의 일부(예를 들어, 사용자의 손)의 포즈(예를 들어, 위치 및/또는 방향)를 검출하는 센서(예를 들어, 포즈 센서)를 포함하거나 이에 (통신 가능하게) 연결될 수 있다. 센서는 포즈의 변화율(1차 미분(들), 예를 들어, 속도, 각속도/속도들, 요/롤/피치 레이트(들))을 더 검출할 수 있다. 센서는 또한 변화율의 변화율(2차 미분(들), 예를 들어, 가속도, 각가속도(들))을 더 포함할 수 있다. 센서에 의한 센서 데이터 출력은 포즈 정보

로 지칭될 수 있다. 일반적으로 포즈 정보는 사용자 또는 사용자의 일부(예를 들어, 사용자의 머리)의 위치 및/또는 방향(포즈)을 나타냄을 이해할 것이다. 또한, 포즈 정보는 포즈의 하나 이상의 변화율(1차 미분)을 나타낼 수 있다. 추가로 더, 포즈 정보는 하나 이상의 변화율의 변화율(2차 미분), 예를 들어, 포즈의 하나 이상의 변화율의 변화율을 나타낼 수 있다. 센서는 예를 들어 컴퓨터 매개 현실 장비(예를 들어, VR/AR/MR 헤드셋/고글)의 일부로서, 또는 사용자가 휴대하는 이동 (컴퓨팅) 디바이스(예를 들어, 스마트폰, 게임 컨트롤러)의 일부로서 사용자 또는 사용자의 관련된 부분(예를 들어, 머리)에 등록하여 배열될 수 있다. 이 경우에 센서는 내장 센서로 지칭될 수 있다. 대안적으로, 센서는 사용자(또는 사용자의 일부)의 포즈를 추적하는 위치 서버(positional server)(예를 들어, 옵티트랙(OptiTrack) 시스템 또는 옵티트랙 타입 시스템)에 의해 제공되거나 구현될 수 있다. 일반적으로, 센서는 사용자(또는 사용자의 일부)의 포즈를 추적하는 추적 시스템의 일부이거나 이에 의해 구현된다. 이러한 위치 서버는 또한 하나 이상의 사용자의 포즈를 추적할 수 있다. 제2 장치는 예를 들어 송신기, 송신기 장치, 서버, 서버 장치 또는 콘텐츠 전달 장치로 지칭될 수 있다. 제1 및 제2 장치 각각은 각각의 메모리에 연결되며 이하에서 설명되는 각각의 작업을 수행하도록 적응(구성)되는 프로세서(또는 프로세서 세트)를 포함할 수 있다. 예를 들어, 상기 프로세서(또는 프로세서 세트)는 이하에서 설명되는 방법(700)의 각 단계를 수행하도록 적응(구성)될 수 있다. 대안적으로, 또는 부가적으로, 상기 프로세서(또는 프로세서 세트)는 이하에서 추가로 설명되는 방법(800), 방법(900), 방법(1000) 중 하나의 각 단계를 수행하도록 적응(구성)될 수 있다.

[0047] S710 단계에서, 사용자(또는 사용자의 일부, 예를 들어 사용자의 머리)의 위치 및/또는 방향을 나타내는 포즈 정보가 획득된다(예를 들어, 결정된다). 이 작업은 센서(예를 들어, 포즈 센서)에 의해 수행될 수 있다. S720 단계에서, 포즈 정보가 제2 장치로 전송된다. S730 단계에서, 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠가 렌더링된다. 즉, 미디어 콘텐츠는 사용자 또는 사용자의 일부의 위치 및/또는 방향에 기반하여 렌더링된다. 렌더링된 미디어 콘텐츠는 사전 렌더링된 미디어 콘텐츠(예를 들어, 사전 렌더링된 오디오 콘텐츠 및/또는 사전 렌더링된 비디오 콘텐츠)로 지칭될 수 있다. 미디어 콘텐츠가 오디오 콘텐츠를 포함하면, 오디오 콘텐츠는 예를 들어 양이 오디오 콘텐츠, B-포맷 오디오 콘텐츠, HOA 오디오 콘텐츠, 채널 기반 오디오 콘텐츠 또는 그 조합으로 렌더링될 수 있다. 일반적으로, 오디오 콘텐츠는 둘 이상의 채널 및/또는 성분으로 렌더링될 수 있다. 미디어 콘텐츠가 비디오 콘텐츠를 포함하면, 예를 들어 비디오 콘텐츠는 타일링(tiling)될 수 있고, 전체 비디오 장면의 관심 영역이 렌더링된 비디오 콘텐츠로 출력될 수 있다. S740 단계에서, 렌더링된 미디어 콘텐츠는 재현을 위해 제1 장치로 전송된다. 단계 S710 및 S720은 제1 장치에서/에 의해, 예를 들어, 센서(예를 들어, 포즈 센서) 및 (제1) 전송 유닛에 의해 각각 수행될 수 있다. 단계 S730 및 S740은 제2 장치에서/에 의해, 예를 들어, 렌더러 및 (제2) 전송 유닛에 의해 수행될 수 있다.

[0048] 오디오 콘텐츠에 대해, 방법(700)은 예를 들어 제1 장치의 일부이거나 이에 연결된 둘 이상의 라우드스피커를 통해 렌더링된 오디오 콘텐츠의 가청(예를 들어, 청각) 표현을 생성하는 단계를 더 포함할 수 있다. 둘 이상의 라우드스피커는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 비디오 콘텐츠에 대해, 방법(700)은 예를 들어 제1 장치의 일부이거나 이에 연결된 디스플레이 디바이스를 통해 렌더링된 비디오 콘텐츠의 시각 표현을 생성하는 단계를 더 포함할 수 있다. 디스플레이 디바이스는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 일반적으로 이러한 표현의 생성은 제1 장치에서/에 의해 수행될 수 있다.

[0049] 위의 방법에 따른 양이 렌더링을 위한 수신기/클라이언트 시스템(200)의 예가 도 2에 개략적으로 도시된다. 시스템은 방법(700)의 제1 장치를 구현할 수 있다. 시스템(200)은, 렌더링된 미디어 콘텐츠(렌더링된 오디오 콘텐츠)의 예로서, 오디오 입력(201)을 수신할 수 있다. 오디오 입력(201)은 예를 들어 양이화된, 비압축 오디오 형태일 수 있다. 수신기 시스템(200)은 (포즈 정보의 예로서) 사용자 움직임 및/또는 사용자 머리 방향에 관한 센서 데이터를 출력할 수 있다. 머리 포즈/센서 데이터(220)는 예를 들어 요, 피치, 롤 및/또는 (x, y, z) 좌표에 관한 정보를 포함할 수 있다. 수신기 시스템(200)은 센서 데이터를 송신기/서버로 출력할 수 있다. 송신기/서버는 방법(700)의 제2 장치를 구현할 수 있다. 수신기 시스템(200)은 오디오 입력(201)의 가청 표현을 더 생성할 수 있다. 예를 들어, 수신기 시스템은 비압축 오디오 입력(201)을 헤드폰 출력으로 출력할 수 있다.

[0050] 이후에 더 자세히 설명될 바와 같이, 도 3, 도 4, 도 5 및 도 6에 도시된 임의의 시스템은 방법(700)을 구현할 수 있다.

[0051] 포즈의 변화와 사용자에게 표시되는 미디어 콘텐츠 표현의 대응하는 적응 사이의 지연을 더 감소시키기 위하여, 제2 장치는 포즈 정보를 예측하여 제1 장치로의 전송 및/또는 인코딩/디코딩(이하에서 설명됨)으로 발생할 수 있는 지연을 예상할 수 있다. 예를 들어, 방법(700)의 S730 단계에서의 미디어 콘텐츠 렌더링은 예측 포즈 정보를 획득(예를 들어, 결정, 계산)하는 것 및 (제1 장치로부터 수신된 포즈 정보에 기반하는 대신) 예측 포즈 정

보에 기반하여 미디어 콘텐츠를 렌더링하는 것을 수반할 수 있다.

[0052] 도 8은 포즈 정보의 예측을 적용하여 미디어 콘텐츠를 처리하는 방법(800)의 예를 개략적으로 도시하는 흐름도이다. 달리 지시하지 않으면, 위의 방법(700)에 관해 이루어진 설명이 또한 여기에서 적용된다.

[0053] S810 단계 및 S820 단계는 각각 방법(700)의 S710 단계 및 S720 단계에 대응한다. S830a 단계에서, S820 단계에서 수신된 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보가 획득(예를 들어, 결정, 계산)된다. 포즈 정보가 포즈의 1차 및/또는 2차 미분을 포함하면, 예측은 이전 포즈 정보에 추가하여 또는 그 대신 상기 1차 및/또는 2차 미분에 기반할 수 있다. 예측 포즈 정보는 미래 타이밍에서 사용자 또는 사용자의 일부(예를 들어, 머리)의 위치 및/또는 방향을 나타내는 미래 타이밍에 대한 포즈 정보일 수 있다. 특정한 구현에서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리되도록 예상되는 타이밍의 추정에 대해 예측될 수 있다. 제1 장치가 재현을 위해 렌더링된 미디어를 처리하도록 예상되는 타이밍의 추정은 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간(지속시간)의 추정을 포함할 수 있다. 대안적으로, 또는 부가적으로, 인코딩/디코딩(예를 들어, 압축/압축해제)이 적용되면(후술됨), 상기 타이밍의 추정은 렌더링된 미디어 콘텐츠를 인코딩/디코딩하는 데 필요한 시간(지속시간)의 추정을 포함할 수 있다. 즉, 예측 포즈 정보는 렌더링된 미디어 콘텐츠의 전송에 필요한 시간 및/또는 렌더링된 미디어 콘텐츠의 인코딩/디코딩에 필요한 시간의 추정에 더 기반하여 획득될 수 있다. 이전 포즈 정보는 이전 타이밍에 제1 장치로부터 수신된 포즈 정보일 수 있다. 이전 포즈 정보의 하나 이상의 항목이 예를 들어 외삽 또는 모델 기반 예측 기술을 통해 예측 포즈 정보를 획득하는 데 사용될 수 있다. 이를 위하여, 이전 포즈 정보의 항목(예를 들어, 소정 수의 항목)이 저장될 수 있다. S830b 단계에서, 렌더링된 미디어 콘텐츠를 획득하기 위해 미디어 콘텐츠가 예측 포즈 정보에 기반하여 렌더링된다. 이 작업은 (S720 단계 또는 S820 단계에서 수신된) 포즈 정보 대신 예측 포즈 정보가 사용되는 점에서 방법(700)의 S730 단계와 다를 수 있지만, 그렇지 않으면 S730 단계와 동일한 방식으로 수행될 수 있다. S840 단계에서, 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치로 전송된다. S810 단계 및 S820 단계는 제1 장치에서/에 의해 수행될 수 있다. S830a 단계, S830b 단계 및 S840 단계는 제2 장치에서/에 의해 수행될 수 있다. S830a 단계는 예를 들어 포즈 예측기에 의해 수행될 수 있다.

[0054] 오디오 콘텐츠에 대하여, 방법(800)은 예를 들어 제1 장치의 일부이거나 이에 연결된 둘 이상의 라우드스피커를 통해 렌더링된 오디오 콘텐츠의 가청(예를 들어, 음향) 표현을 생성하는 단계를 더 포함할 수 있다. 둘 이상의 라우드스피커는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 비디오 콘텐츠에 대하여, 방법(800)은 예를 들어 제1 장치의 일부이거나 이에 연결된 디스플레이 디바이스를 통해 렌더링된 비디오 콘텐츠의 시각 표현을 생성하는 단계를 더 포함할 수 있다. 디스플레이 디바이스는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 일반적으로, 이러한 표현의 생성은 제1 장치에서/에 의해 수행될 수 있다.

[0055] 방법(800)의 변형에서, 예측 포즈 정보가 제1 장치에서 예측될 수 있다. 즉, 제1 장치는 S830a 단계를 참조하여 위에서 설명한 바와 같은 처리를 수행하고, 그 다음 예측 포즈 정보를 제2 장치로 송신할 수 있다. 이 경우 S820 단계가 생략될 수 있음을 이해할 것이다. 제1 장치로부터 예측 포즈 정보를 수신한 후에, 제2 장치는 상술한 방식으로 S830b 단계 및 후속 단계 처리를 계속할 수 있다.

[0056] 이후에 더 자세히 설명될 바와 같이, 도 3, 도 4, 도 5 및 도 6에 도시된 임의의 시스템은 방법(800) 또는 방법(800)의 변형을 구현할 수 있다.

[0057] 미디어 콘텐츠 렌더링을 위한 위의 포즈 정보 예측은 전송 및/또는 인코딩/디코딩에 의해 일어나는 지연을 "감출" 수 있게 하여, 사용자 움직임과 렌더링된 미디어 콘텐츠의 표시 사이에 양호한 정렬이 달성될 수 있다. 따라서, 사용자가 멀미의 영향을 받을 위험이 줄어들거나 회피될 수 있으며, 사용자의 몰입형 컴퓨터 매개 현실 경험이 개선될 수 있다. 방법(800)의 경우, 움직임과 렌더링된 미디어 콘텐츠의 표시 사이의 정렬의 개선은 서버/송신기 측에 의해 수행되는 프로세스에 의해, 즉, 포즈 정보를 예측하고 미디어 콘텐츠를 렌더링하기 위하여 수신기/재생 측으로부터 수신한 포즈 정보 대신 예측 포즈 정보를 사용함에 의하여 달성된다. 그러나, 특정한 조건 하에서 움직임과 렌더링된 미디어 콘텐츠의 표시 사이의 정렬의 이러한 개선을 수신기 또는 재생 측에서 수행되는 조치에 의해 달성하는 것이 바람직할 수 있다.

[0058] 도 9는 위의 고려사항에 따른, 즉 움직임과 렌더링된 미디어 콘텐츠의 표시 사이의 정렬을 수신기/재생 측에서 수행되는 조치에 의해 개선하는 미디어 콘텐츠를 처리하는 방법(900)의 예를 개략적으로 도시하는 흐름도이다.

[0059] S910 단계, S920 단계, S930 단계 및 S940 단계는 각각 방법(700)의 S710 단계 내지 S740 단계에 대응한다. S950 단계에서, 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보(예를 들어, 제1 장치로부터 수신된 포즈 정

보)가 제1 장치로 전송된다. 상기 포즈 정보는 렌더링된 미디어 콘텐츠와 함께, 예를 들어, 렌더링된 미디어 콘텐츠와 연관되어 전송될 수 있다. S960 단계에서, 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보를 현재 포즈 정보와 비교한다. 현재 포즈 정보는 렌더링된 미디어 콘텐츠를 재현(예를 들어, 재생)하는 시간에 획득된 포즈 정보일 수 있다. 현재 포즈 정보는, 다른 (이후의) 시간에서이지만, S710 단계를 참조하여 상술된 방식으로 획득될 수 있다. S970 단계에서, 렌더링된 미디어 콘텐츠는 비교 결과에 기반하여 업데이트된다. 예를 들어, 렌더링된 미디어 콘텐츠는 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보와 현재 포즈 정보 사이의 차이에 기반하여 업데이트될 수 있다. 상기 업데이트는 렌더링된 미디어 콘텐츠의 외삽을 수반할 수 있다. 이러한 업데이트의 비제한적인 예가 도 3을 참조하여 이하에서 설명될 것이다. S910 단계, S920 단계, S960 단계 및 S970 단계는 제1 장치에서/에 의해 수행될 수 있다. S930 단계, S940 단계 및 S950 단계는 제2 장치에서/에 의해 수행될 수 있다.

[0060] 특정한 구현에서, S950 단계는 생략될 수 있다. 즉, 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보가 제1 장치로 전송되지 않을 수 있다. 이 경우, S920 단계에서 제2 장치로 송신된 포즈 정보가, S960 단계에서, 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보로 지칭될 수 있다.

[0061] 또한, 특정한 구현에서, 방법(900)이 포즈 정보의 변화에 응답하여 (예를 들어, 사용자의 포즈 또는 사용자 머리의 포즈의 변화에 응답하여) 렌더링된 미디어 콘텐츠가 변화하는 방법인 기울기 정보를 결정하는 것을 포함할 수 있다. 그러면, 방법(900)은 추가로 또한 기울기 정보를 제1 장치로 전송하는 것을 포함할 수 있다. 예를 들어, 기울기 정보는 렌더링된 미디어 콘텐츠 및 선택적으로 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보와 함께(예를 들어, 이와 연관되어) 제1 장치로 전송될 수 있다. 이들 추가적인 단계는 제2 장치에서 수행될 수 있다. 오디오 콘텐츠에 대하여, 기울기 정보는 사용자 또는 사용자의 일부의 병진 및/또는 회전에 응답하는 (예를 들어, 각 채널 또는 각 성분의) 부대역 에너지 레벨의 변화를 나타낼 수 있다. 기울기 정보는 S970 단계에서 렌더링된 미디어 콘텐츠를 업데이트/조정하기 위하여 사용될 수 있다. 예를 들어, 렌더링된 오디오 콘텐츠의 부대역 에너지 레벨이 기울기 정보 및 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보와 현재 포즈 정보 사이의 차이에 기반하여 조정될 수 있다. 일반적으로, 렌더링된 미디어 콘텐츠는 포즈의 차이 및 포즈의 변화에 응답하는 렌더링된 미디어 콘텐츠의 변화를 나타내는 기울기에 기반하여 업데이트/조정될 수 있다.

[0062] 오디오 콘텐츠에 대하여, 방법(900)은 예를 들어 제1 장치의 일부이거나 이에 연결된 둘 이상의 라우드스피커를 통해 렌더링된 오디오 콘텐츠의 가청(예를 들어, 음향) 표현을 생성하는 단계를 더 포함할 수 있다. 둘 이상의 라우드스피커는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 비디오 콘텐츠에 대하여, 방법(900)은 예를 들어 제1 장치의 일부이거나 이에 연결된 디스플레이 디바이스를 통해 렌더링된 비디오 콘텐츠의 시각 표현을 생성하는 단계를 더 포함할 수 있다. 디스플레이 디바이스는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 일반적으로, 이러한 표현의 생성은 제1 장치에서/에 의해 수행될 수 있다.

[0063] 이후에 더 자세히 설명될 바와 같이, 도 3, 도 4, 도 5 및 도 6에 도시된 임의의 시스템은 방법(900)을 구현할 수 있다.

[0064] 사용자 움직임과 렌더링된 미디어 콘텐츠의 표시 사이의 정렬을 더 개선하기 위하여, 서버/송신기 측의 포즈 정보 예측과 수신기/재생 측의 렌더링된 미디어 콘텐츠 업데이트가 결합될 수 있다.

[0065] 도 10은 위의 고려사항에 따른, 즉 움직임과 렌더링된 미디어 콘텐츠의 표시 사이의 정렬을 서버/송신기 측에서 수행되는 조치(measure)뿐 아니라 수신기/재생 측에서 수행되는 조치에 의해 개선하는 미디어 콘텐츠를 처리하는 방법(1000)의 예를 개략적으로 도시하는 흐름도이다.

[0066] S1010 단계, S1020 단계 및 S1040 단계는 각각 방법(700)의 S710 단계, S720 단계 및 S740 단계에 대응한다. S1030a 단계 및 S1030b 단계는 각각 방법(800)의 S830 단계 및 S830b 단계에 대응한다. S1050 단계에서, 예측 포즈 정보(즉, 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보)가 제1 장치로 전송된다. 예측 포즈 정보는 렌더링된 미디어 콘텐츠와 함께, 예를 들어, 렌더링된 미디어 콘텐츠와 연관되어 전송될 수 있다. S1060 단계에서, 예측 포즈 정보를 실제/현재 포즈 정보와 비교한다. 실제 포즈 정보는 렌더링된 미디어 콘텐츠를 재현(예를 들어, 재생)하는 시간에 획득되는 포즈 정보일 수 있다. 실제 포즈 정보는 다른 (이후의) 시간에서이지만, S710 단계를 참조하여 상술된 방식으로 획득될 수 있다. S1070 단계에서, 렌더링된 미디어 콘텐츠는 비교 결과에 기반하여 업데이트된다. 예를 들어, 렌더링된 미디어 콘텐츠는 예측 포즈 정보와 실제 포즈 정보 사이의 차이에 기반하여 업데이트될 수 있다. 일반적으로, 업데이트는 방법(900)의 S970 단계에서와 동일한 방식으로 수행될 수 있다. S1010 단계, S1020 단계, S1060 단계 및 S1070 단계는 제1 장치에서/에 의해 수행될 수 있다. S1030a 단계, S1030b 단계, S1040 단계 및 S1050 단계는 제2 장치에서/에 의해 수행될 수 있다.

[0067] 특정한 구현에서, 방법(1000)이 포즈 정보의 변화에 응답하여 (예를 들어, 사용자의 포즈 또는 사용자 머리의 포즈의 변화에 응답하여) 렌더링된 미디어 콘텐츠가 변화하는 방법인 기울기 정보를 결정하는 것을 포함할 수 있다. 그러면, 방법(1000)은 추가로 또한 기울기 정보를 제1 장치로 전송하는 것을 포함할 수 있다. 예를 들어, 기울기 정보는 렌더링된 미디어 콘텐츠 및 선택적으로 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보와 함께(예를 들어, 이와 연관되어) 제1 장치로 전송될 수 있다. 이들 추가적인 단계는 제2 장치에서 수행될 수 있다. 오디오 콘텐츠에 대하여, 기울기 정보는 사용자 또는 사용자의 일부의 병진 및/또는 회전에 응답하는 (예를 들어, 각 채널 또는 각 성분의) 부대역 에너지 레벨의 변화를 나타낼 수 있다. 기울기 정보는 S1070 단계에서 렌더링된 미디어 콘텐츠를 업데이트/조정하기 위하여 사용될 수 있다. 예를 들어, 렌더링된 오디오 콘텐츠의 부대역 에너지 레벨이 기울기 정보 및 미디어 콘텐츠를 렌더링하기 위해 사용된 포즈 정보와 현재 포즈 정보 사이의 차이에 기반하여 조정될 수 있다. 일반적으로, 렌더링된 미디어 콘텐츠는 포즈의 차이 및 포즈의 변화에 응답하는 렌더링된 미디어 콘텐츠의 변화를 나타내는 기울기에 기반하여 업데이트/조정될 수 있다.

[0068] 오디오 콘텐츠에 대하여, 방법(1000)은 예를 들어 제1 장치의 일부이거나 이에 연결된 둘 이상의 라우드스피커를 통해 렌더링된 오디오 콘텐츠의 가청(예를 들어, 음향) 표현을 생성하는 단계를 더 포함할 수 있다. 둘 이상의 라우드스피커는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 비디오 콘텐츠에 대하여, 방법(1000)은 예를 들어 제1 장치의 일부이거나 이에 연결된 디스플레이 디바이스를 통해 렌더링된 비디오 콘텐츠의 시각 표현을 생성하는 단계를 더 포함할 수 있다. 디스플레이 디바이스는 예를 들어 컴퓨터 매개 현실 장비의 일부일 수 있다. 일반적으로, 이러한 표현의 생성은 제1 장치에서/에 의해 수행될 수 있다.

[0069] 방법(1000)의 변형에서, 예측 포즈 정보가 제1 장치에서 예측될 수 있다. 즉, 제1 장치는 S1030a 단계를 참조하여 위에서 설명한 바와 같은 처리를 수행하고, 그 다음 예측 포즈 정보를 제2 장치로 송신할 수 있다. 이 경우 S1020 단계가 생략될 수 있음을 이해할 것이다. 제1 장치로부터 예측 포즈 정보를 수신한 후에, 제2 장치는 예측 포즈 정보를 이용하여, S1030b 단계를 참조하여 상술한 방식으로 미디어 콘텐츠를 렌더링하고, S1040 단계를 참조하여 상술한 방식으로 렌더링된 미디어 콘텐츠를 제1 장치로 전송할 수 있다. 이 경우 S1050 단계는 생략될 수 있다. 렌더링된 미디어 콘텐츠를 수신한 후에, 제1 장치는 상술한 방식으로 S1060 단계 및 S1070 단계를 수행할 수 있다. 특히, 이 경우 포즈 정보 예측이 제1 장치에서 수행되므로, 제1 장치는 제2 장치로부터 예측 포즈 정보를 수신할 필요가 없다.

[0070] 이후에 더 자세히 설명될 바와 같이, 도 3, 도 4, 도 5 및 도 6에 도시된 임의의 시스템은 방법(1000) 또는 방법(1000)의 변형을 구현할 수 있다.

[0071] 위 방법 중 임의의 하나에서, 렌더링된 미디어 콘텐츠는 비압축 형태로 제1 장치로 전송될 수 있다. 이는 제2 장치에서의 사전 렌더링에 의해 가능하게 되어, 전체 미디어 콘텐츠(예를 들어, 오디오/비디오 장면의 완전한 표현)의 전송이 필수적이지 않게 된다. 렌더링된 미디어 콘텐츠를 비압축 형태로 전송하는 것은 왕복 지연을 줄이는 데 기여하는데, 이는 압축/압축해제에 통상 소요되는 시간이 절약되기 때문이다. 한편, 대역폭 제한에 의해 요구되는 경우 렌더링된 미디어 콘텐츠가 제1 장치로 전송되기 전에 인코딩(압축)될 수 있다. 이 경우, 위에서 언급한 바와 같이, 예측 포즈 정보를 획득할 때 인코딩/디코딩(예를 들어, 압축/압축해제)에 필요한 시간이 고려될 수 있다.

[0072] 또한, 오디오 콘텐츠에 대하여 위의 방법 중 임의의 하나가 제1 장치가 위치한 환경의 음향 특성을 나타내는 환경 정보를 제2 장치로 전송하는 것을 더 포함할 수 있다. 환경 정보는 룸 특성 및/또는 양이 룸 임펄스 응답(BRIR) 함수를 포함할 수 있다. 이 단계는 제1 장치에서/에 의해, 예를 들어, 셋업 시에 수행될 수 있다. 그런 다음, 오디오 콘텐츠는 환경 정보에 추가로 기반하여 렌더링될 수 있다. 대안적으로, 또는 부가적으로, 위의 방법 중 임의의 하나가 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보를 제2 장치로 전송하는 것을 더 포함할 수 있다. 형태는 형상 또는 크기, 예를 들어, 사용자 머리의 형상 또는 크기를 포함하거나 이에 대응할 수 있다. 형태 정보는 머리 전달 함수(HRTF)를 포함할 수 있다. 렌더링은 양이 렌더링일 수 있다. 이 단계는 제1 장치에서/에 의해, 예를 들어, 셋업 시에 수행될 수 있다. 그런 다음, 오디오 콘텐츠는 형태 정보에 추가로 기반하여 렌더링될 수 있다.

[0073] 도 3은 서버/송신기(300) 및 클라이언트/수신기(350)를 포함하는 예시적인 시스템의 추가적인 세부사항을 도시한다. 위에서 언급한 바와 같이, 이 시스템은 방법(700, 800, 900 및 1000) 중 임의의 것을 구현할 수 있다. (예를 들어, 제2 장치를 구현하는) 서버/송신기(300)는 렌더러(320)(예를 들어, 오디오 렌더러) 및 인코더(330)를 포함할 수 있다. (예를 들어, 제1 장치를 구현하는) 클라이언트/수신기(350)는 현재 시점 t0에서 현재 포즈(예를 들어, 머리 포즈) P(t0)를 서버/송신기(300)로 송신할 수 있다. 현재 포즈 P(t0)는 또한 현재 포즈

P(t0)가 생성된 시간을 특정하는 타임스탬프 t0 자체를 포함할 수 있다. 포즈 P(t0)는 포즈(350) 블록에 의해 결정되고 송신될 수 있다.

[0074] (예를 들어, 제2 장치를 구현하는) 서버/송신기(300)는 위치 예측기(310)를 더 포함할 수 있다. 서버/송신기(300)는 시점 t1에 사용자 위치 및 현재 포즈(머리 방향에 대응함) P(t0)를 수신할 수 있으며, t1 > t0이다. 수신된 현재 포즈 P(t0) 및 t0 자체가 위치 P(t1)를 예측하기 위해 위치 예측기(310)에 의해 사용될 수 있다. 위치 예측기(310)는 위치 P(t1) 예측을 위하여 이전에 수신된 포즈 P(tn) 및 tn을 고려할 수 있으며, n은 0 내지 무한일 수 있다(시간상 빠른 시점으로부터의 포즈 및 타임스탬프 값). 위치 P(t1)는 포즈 P(t0)와 유사할 수 있다. 위치 P(t1)는 오디오 렌더러(320)에 의해 사용되어 시점 t1에서 오디오 장면을 렌더링하고 이에 따라 렌더링된 오디오 데이터 R(t1)(340)를 결정할 수 있다. 렌더링된 오디오 데이터 R(t1)(340)는 오디오 인코더(330)를 사용하여 인코딩되어 오디오 데이터 A(t1)를 결정할 수 있다. 서버/송신기(300)는 오디오 데이터 A(t1) 및 위치 P(t1)를 클라이언트/수신기(350)로 송신할 수 있다. 위치 P(t1)는 오디오 비트스트림의 일부로서 인코딩될 수 있다. 클라이언트/수신기(350)는 시점 t2에서 서버/송신기(300)로부터 오디오 데이터 A(t1) 및 위치 P(t1)를 (예를 들어, 메타데이터의 형태로) 수신할 수 있으며, t2 > t1이다. 클라이언트/수신기(350)는 오디오 데이터 A(t1) 및 위치 P(t1)를 오디오 디코더(351)에서 수신할 수 있으며 이는 비압축된 오디오 U(t1)을 결정할 수 있다. 머리 포즈/센서 데이터(352) 블록은 시점 t2에서의 포즈 P(t2)를 결정할 수 있다. 오디오 외삽기(353)는 수신된 P(t1)를 사용하여 시점 t2에서의 포즈 P(t2)에서 포즈 P(t1)을 차감함으로써 포즈 차이 DeltaP를 계산할 수 있다. DeltaP는 오디오 외삽기(353)에 의해 사용되어 출력(390) 전의 비압축 오디오 U(t1)를 적응/외삽할 수 있다. 클라이언트/수신기(350)는 오디오 콘텐츠가 FOA이고 모션이 요, 피치 및/또는 롤 움직임으로 제한되는 경우 외삽의 일부로 국소적 회전을 적용할 수 있다. 클라이언트/수신기(350)는 또한 오디오 콘텐츠가 사전 렌더링된 양이 콘텐츠 또는 사전 렌더링된 채널 기반 콘텐츠인 경우 외삽의 일부로 블라인드 업믹싱(blind upmixing)을 적용할 수 있다.

[0075] 위치 P(t1)를 예측하는 대신, 클라이언트/수신기(350)가 오디오 데이터를 수신 또는 처리하도록 예상되는 시점 t2'에 대해 위치 P(t2')가 예측될 수 있다. 시점 t2'는 시점 t1으로부터 출발하여, 오디오 데이터 전송 및/또는 인코딩/디코딩에 필수적인 시간(지속시간)을 고려하여 추정될 수 있다. 그러면 위의 P(t1), R(t1), A(t1) 및 U(t1)은 각각 P(t2'), R(t2'), A(t2') 및 U(t2')로 대체되어야 할 것이다. 임의의 상술한 요소가 각 장치의 프로세서(또는 프로세서 세트)에 의해 구현될 수 있다.

[0076] MPEG-H 3D 오디오 (ISO/IEC 23008-3) 및/또는 MPEG 표준의 미래 버전의 다음 신택스(syntax)가 P(t) 3자유도(3DoF(Degrees of Freedom)) 데이터를 전송하는 데 사용될 수 있다.

표 1

[0077]

신택스	비트 수	니모닉(Mnemonic)
mpegh3daSceneDisplacementData()		
{		
sd_yaw;	9	uimsbf
sd_pitch;	9	uimsbf
sd_roll;	9	uimsbf
}		

[0078] 시맨틱은 MPEG-H 3D 오디오 (ISO/IEC 23008-3) 및/또는 MPEG 표준의 미래 버전에 따라 정의될 수 있다.

[0079] 6DoF 데이터 및 타임스탬프 전송용 전체 신택스는 다음과 같을 수 있다.

표 2

신택스	비트수
head_pose() {	

if (b_yaw) {
yaw;
}
	1
	9

if (b_pitch) {
pitch;
}
	1
	9

if (b_roll) {
roll;
}
	1
	9

if (b_x_pos) {
x_pos;
}
	1
	9

if (b_y_pos) {
y_pos;
}
	1
	9

if (b_z_pos) {
z_pos;
}
	1
	9

if (b_timestamp) {
timestamp;
}
	1
	variable

if (PosDerivatives) {
d_x_pos
d_y_pos
d_z_pos
dd_x_pos
dd_y_pos
dd_z_pos
}
	1
	9
	9
	9
	9
	9
	9

if (OrientDerivatives) {
d_roll
d_pitch
d_yaw
dd_roll
dd_pitch
dd_yaw
}
	1
	9
	9
	9
	9
	9
	9

}	

[0080]

[0081] 시맨틱은 MPEG-H 3D 오디오 (ISO/IEC 23008-3) 및/또는 MPEG 표준의 미래 버전에 따라 정의될 수 있다.

[0082] 도 4는 송신기(400) 및 수신기(450)를 포함하는 예시적인 시스템을 도시한다. 도 4에 도시된 시스템은 도 1 내지 도 3에 도시된 시스템의 전체 또는 일부 양상을 포함할 수 있다. 특히, 시스템은 임의의 상술한 방법(700, 800, 900 및 1000) 또는 모두를 구현할 수 있다. (예를 들어, 제2 장치를 구현하는) 송신기/서버(400)는 (미디어 콘텐츠의 예로서) 완전한 오디오 장면(예를 들어, FOA/HOA 또는 객체 기반 장면)(401) 및 완전한 비디오 장면(예를 들어, 360° 비디오)(402)을 수신할 수 있다. 오디오 장면(401)은 오디오 인코더(410)에 의해 처리되어 완전한 오디오 장면 전달(441)을 결정할 수 있다. 완전한 오디오 장면(441)은 전체 오디오 장면 및/또는 부수하는 메타데이터(예컨대 오디오 객체 위치, 방향 등)로 구성될 수 있다. 완전한 비디오(402)는 콘텐츠 전달 선택(420)에 의해 처리될 수 있다. 완전한 비디오(420)는 관심 영역과 같은 상이한 부분으로 나누어지고 이에 따라 콘텐츠 선택(420)에 의해 "타일링"되어(360° 비디오가 타일로 나누어질 수 있음) 타일(402a)을 결정할 수 있다. 콘텐츠 전달 및 선택(420)은 도 3에서 위치 예측기(310)의 출력으로 설명되는 예측 위치 P(t1)(또는 예측 위치 P(t2'))를 사용할 수 있거나 또는 변경되지 않은 머리 포즈/센서 데이터(454)를 사용할 수 있다. 예를 들

어, 완전한 360° 비디오(402) 중에서 타일(402a)이 수신기(450)로부터 수신한 센서 데이터(454)에 기반하여 콘텐츠 전달 선택(420)에서 선택될 수 있다. 이 선택은 비디오 콘텐츠의 렌더링으로 지칭될 수 있다. 비디오 인코더(430)는 타일(402a)을 인코딩하여 (예를 들어, 제1 장치를 구현하는) 클라이언트/수신기(450)로 전송될 수 있는 관심 영역 비디오(442)를 출력한다. 수신기(450)는 관심 영역 비디오(442)를 수신할 수 있는 비디오 디코더(452)를 포함할 수 있다. 비디오 디코더(452)는 관심 영역(442)을 사용하여 비디오를 디코딩하고 이를 비디오 출력(492)으로 출력한다. 완전한 오디오 장면(441)이 콘텐츠를 디코딩할 수 있고 디코딩된 오디오 장면을 오디오 렌더러(453)로 제공할 수 있는 오디오 디코더(451)에 의해 수신될 수 있다. 오디오 디코더(451)는 비압축 오디오 및 메타데이터(455)(디코딩된 오디오 장면에 대응할 수 있음)를 오디오 렌더러(453)에 제공할 수 있다. 오디오 렌더러(453)는 센서 데이터(454)에 기반하여 디코딩된 오디오를 렌더링할 수 있고 오디오 출력(491)을 출력할 수 있다. 센서 데이터(454)는 사용자 움직임 및/또는 사용자 머리 방향을 검출할 수 있는 센서(예를 들어 자이로스코프 기반 센서)로부터 수신될 수 있다. 그런 다음 이는 또한 완전한 오디오 장면(441)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 오디오 렌더러(453)로 제공되고 완전한 비디오 장면(402)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 콘텐츠 전달 선택(420)으로 제공될 수 있다. 특히, 도 4의 예시적인 시스템에서 오디오 콘텐츠가 수신기/재생 측에서 렌더링되는 반면, 비디오 콘텐츠는 서버/송신기 측에서 렌더링된다(즉, 수신기/재생 측에서의 재생 준비가 완료된 비디오 콘텐츠가 서버/송신기 측에서 생성된다). 상술한 임의의 요소는 각 장치의 프로세서(또는 프로세서 세트)에 의해 구현될 수 있다.

[0083] 도 5는 송신기(500) 및 수신기(550)를 포함하는 예시적인 시스템을 도시한다. 시스템은 임의의 상술한 방법(700, 800, 900 및 1000) 또는 모두를 구현할 수 있다. (예를 들어, 제2 장치를 구현하는) 송신기/서버(500)는 (미디어 콘텐츠의 예로서) 완전한 오디오 장면(예를 들어, HOA 또는 객체 기반 장면)(501) 및 완전한 비디오 장면(예를 들어, 360° 비디오)(502)을 수신할 수 있다. 오디오 장면(501)은 오디오 렌더러(510)에 의해 처리되어 양이 오디오 데이터(541)를 결정할 수 있다. 오디오 렌더러(510)는 센서 데이터(545)를 고려하여 양이 오디오 데이터(541)를 결정할 수 있다. 센서 데이터(545)는 요, 피치, 롤, x, y, z 정보를 포함할 수 있다. 양이 오디오 데이터(541)는 비압축, 무손실 압축 또는 손실 저 대기시간 압축(lossy low-latency compressed)일 수 있다. 예를 들어, 양이 오디오 데이터(551)는 (예를 들어, 제1 장치를 구현하는) 수신기(550)에 의해 수신되고 양이 오디오 출력(591)으로 제공되는 비압축 오디오(580)일 수 있다. 완전한 비디오(502)는 콘텐츠 전달 선택(520)에 의해 처리될 수 있다. 완전한 비디오(502)는 관심 영역과 같은 상이한 부분으로 나누어지고 이에 따라 콘텐츠 선택(520)에 의해 "타일링"되어(360° 비디오가 타일로 나누어질 수 있음) 타일(502a)을 결정할 수 있다. 완전한 360° 비디오(502) 중에서 타일(502a)이 수신기(550)로부터 수신한 센서 데이터(545)에 기반하여 콘텐츠 전달 선택(520)에서 선택될 수 있다. 이 선택은 비디오 콘텐츠의 렌더링으로 지칭될 수 있다. 비디오 인코더(530)는 타일(502a)을 인코딩하여 클라이언트/수신기(550)로 전송될 수 있는 관심 영역 비디오(542)를 출력한다. 수신기(550)는 관심 영역 비디오(542)를 수신할 수 있는 비디오 디코더(552)를 포함할 수 있다. 비디오 디코더(552)는 관심 영역(542)을 사용하여 비디오를 디코딩하고 이를 비디오 출력(592)으로 출력한다. 센서 데이터(545)는 사용자 움직임 및/또는 사용자 머리 방향을 검출할 수 있는 센서(예를 들어 자이로스코프 기반 센서)로부터 수신될 수 있다. 그런 다음 이는 또한 완전한 비디오 장면(502)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 콘텐츠 전달 선택(520)으로 제공될 수 있다. 그런 다음 이는 또한 완전한 오디오 장면(501)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 콘텐츠 오디오 렌더러(510)로 제공될 수 있다. 상술한 임의의 요소는 각 장치의 프로세서(또는 프로세서 세트)에 의해 구현될 수 있다.

[0084] 도 6은 송신기(600) 및 수신기(650)를 포함하는 예시적인 시스템을 도시한다. 시스템은 임의의 상술한 방법(700, 800, 900 및 1000) 또는 모두를 구현할 수 있다. (예를 들어, 제2 장치를 구현하는) 송신기/서버(600)는 (미디어 콘텐츠의 예로서) 완전한 오디오 장면(예를 들어, HOA 또는 객체 기반 장면)(601) 및 완전한 비디오 장면(예를 들어, 360° 비디오)(602)을 수신할 수 있다. 오디오 장면(601)은 오디오 렌더러(610)에 의해 처리될 수 있고 그런 다음 오디오 렌더러(610)의 출력이 저지연 오디오 인코더(660)에 의해 처리될 수 있다. 오디오 렌더러(610)는 센서 데이터(645)를 고려할 수 있다. 저지연 오디오 인코더(660)는 양이 오디오 데이터(641)를 출력할 수 있으며, 이는 그런 다음 (예를 들어, 제1 장치를 구현하는) 수신기(650)로 송신될 수 있다. 양이 오디오 데이터(641)는 수신기(650)에서 저지연 오디오 디코더(670)에 의해 수신될 수 있으며 이는 양이 오디오 데이터(641)를 비압축 오디오(680)로 변환한다. 비압축 오디오(680)는 이어서 양이 오디오 출력(691)으로 제공될 수 있다. 완전한 비디오(602)는 콘텐츠 전달 선택(620)에 의해 처리될 수 있다. 완전한 비디오(602)는 관심 영역과 같은 상이한 부분으로 나누어지고 이에 따라 콘텐츠 선택(620)에서 "타일링"되어(360° 비디오가 타일로 나누어질 수 있음) 수신기(650)로부터 수신된 센서 데이터(645)에 기반하여 콘텐츠 전달 선택(620)에서 선택될 수 있는 타일을 결정한다. 이 선택은 비디오 콘텐츠의 렌더링으로 지칭될 수 있다. 비디오 인코더(630)는 타일 및/또

는 비디오를 인코딩하여 클라이언트/수신기(650)로 전송될 수 있는 관심 영역 비디오(642)를 출력한다. 수신기(650)는 관심 영역 비디오(642)를 수신할 수 있는 비디오 디코더(652)를 포함할 수 있다. 비디오 디코더(652)는 관심 영역(642)을 사용하여 비디오를 디코딩하고 이를 비디오 출력(692)으로 출력한다. 센서 데이터(645)는 사용자 움직임 및/또는 사용자 머리 방향을 검출할 수 있는 센서(예를 들어 자이로스코프 기반 센서)로부터 수신될 수 있다. 그런 다음 이는 또한 완전한 비디오 장면(602)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 콘텐츠 전달 선택(620)으로 제공될 수 있다. 그런 다음 이는 또한 완전한 오디오 장면(601)을 사용자의 현재 머리 방향 및/또는 위치에 적응시키기 위하여 콘텐츠 오디오 렌더러(610)로 제공될 수 있다. 상술한 임의의 요소는 각 장치의 프로세서(또는 프로세서 세트)에 의해 구현될 수 있다.

[0085] 통상적으로, 송신기(S)로부터 수신기(R)로 전송되는 (미디어 콘텐츠의 비제한적인 예로서) 오디오는 도 1 및 도 4에 나타난 수신기에서 렌더링된다. 수신기 측의 유연성을 최대화하기 위하여, 예를 들어 수신 측에서 적응적으로 렌더링될 수 있는 객체 또는 HOA와 같은 오디오 장면의 복소(complex) 표현을 전송하여, 로컬 청취자의 시점/포즈와 일치시키는 것이 가능하다. 그러나, 이러한 표현을 인코딩하기 위해서는 큰 대기시간이 필요하여 이들 접근이 통신 또는 대화형 애플리케이션에 사용되지 못하게 한다.

[0086] 본 개시는 수신기 내에서 언급된 대기시간을 줄이고 및/또는 계산 복잡도를 줄이는 방법, 시스템 및 장치를 제공한다. 수신기로부터 송신기로 전송되는 사용자 위치 및 방향은 서버/송신기가 수신기 현재 포즈/시점과 밀접하게 일치하는 더 작고, 사전 렌더링된 버전의 콘텐츠를 계산할 수 있게 한다. 그러나, 송신기로부터 수신기로의 전송 대기시간은 로컬 수신 포즈와 서버 상에서 렌더링이 계산된 포즈 사이에 불일치가 도입되도록 할 수 있다. 본 개시는 송신기가 렌더링이 수행된 위치를 신호하여 수신기가 렌더링된 신호를 그 현재의 로컬 포즈에 외삽하도록 허용한다. 또한, 송신기는 사전 렌더링된, 비압축 또는 손실 압축된 오디오 장면 표현을 송신할 수 있어 시스템 내에서 인코딩 및 디코딩 대기시간을 제거한다. 송신기는 예를 들어 양이 스테레오, FOA 또는 HOA 표현을 향한 렌더링 알고리즘을 수행한다. 렌더링 알고리즘은 두 채널을 향한 오디오 객체(예를 들어, 사전 렌더링된 양이화된 콘텐츠)와 같은 오디오 데이터를 출력 채널로 렌더링한다. 그런 다음 채널은, 특히 (예를 들어, 시스템의 대역폭에 따라) 압축이 필요하면 인코딩되어 인코디된 오디오 데이터 비트스트림을 출력한다. 신호는 클라이언트 또는 수신기로 전송될 수 있고 이는 헤드폰 또는 스테레오 라우드스피커 시스템을 통해 출력될 수 있다.

[0087] 사용자의 머리의 물리적 특성에 적응되기 위하여 양이화된 재생이 필요할 때, 수신기는 사용자 머리의 특성에 대응하는 머리 전달 함수(HRTF)를 전송할 수 있다. 수신기는 또한 재현이 의도되는 방에 대응하는 양이 룸 임펄스 응답(BRIR) 함수를 전송할 수 있다. 이 정보는 전송 셋업 동안 전송될 수 있다.

[0088] 본 개시의 실시예는 적어도 다음의 이점을 제공한다.

[0089] · 양이화된(스테레오) 데이터만이 전송되면, 무손실 오디오 데이터 전송의 전송 비트율이 완전한 압축 오디오 장면의 비트율과 비슷하거나 유사할 수 있다.

[0090] · 비압축 또는 무손실 오디오 스트림의 전송은 인코딩 및 디코딩을 위한 대기시간을 제거 또는 감소시킨다. 예를 들어, 인코딩/디코딩으로 인한 대기시간이 0으로 감소될 수 있으며, 이는 모션 투 이어 대기시간의 전반적인 감소를 가져온다.

[0091] · 오디오 데이터의 압축이 없을 때 수신기는 디코딩 또는 렌더링 없이 단지 오디오를 출력하기만 한다. 이는 디코딩이 수행될 필요가 없으며 및/또는 송신기 측에서 렌더링이 이미 완료되었으므로 수신기에서의 계산 복잡도를 감소시킨다.

[0092] · 높은 비트율에서 최소 대기시간 및 최소 수신기 계산 복잡도와 높은 대기시간 및 높은 수신기 계산 복잡도에서 최소 비트율 사이에서 상이한 절충안을 취할 수 있다. 예를 들어:

[0093] o 비압축 데이터를 전송하기에 충분한 대역폭이면 최소 대기시간과 계산 복잡도를 위해 비압축 데이터를 전송

[0094] o 비압축 데이터를 전송하기에 대역폭이 충분하지 않으면, 최소 대기시간과 약간 높은 계산 복잡도로 무손실 압축 데이터를 전송

[0095] o 대역폭이 제한적이면, 낮은 대기시간과 더 높은 계산 복잡도로 손실, 저지연 압축 데이터를 전송

[0096] 위의 내용은 R과 S가 동시에 수신기와 송신기인 경우 R과 S 사이의 양방향 통신에도 적용된다.

[0097] 표 3은 이러한 절충안을 예시하는 시스템 비교의 예를 나타낸다.

표 3

[0098]

코덱	비트율 (전송 품질)	인코딩 대기시간	디코딩 대기시간	전송 대기시간	전체 대기시간	수신기 계산 복잡도	출력 유연성
3D 오디오 코덱	800kbps - 1200kbps	~120ms	~40ms	~5ms	~165ms	높음	높음
사전 렌더링된 무손실 코딩	768kbps	~5ms	~0ms	2*~5ms	~15ms	낮음	낮음

[0099] 특정 문맥에서, 본 개시의 실시예는 콘텐츠 외삽에 기반하여 전송 대기시간을 숨기는 것을 수반할 수 있다. 전체 대기시간(예를 들어, 전송 대기시간)이 너무 높으면(일반적으로 20msec 보다 높은), 다음 업데이트된 콘텐츠 프레임의 전달을 기다리는 동안, 인코딩 포맷 및/또는 재생 시스템은 수신기의 로컬 포즈(위치 및 방향)와 일치하도록 콘텐츠를 외삽하는 수단을 제공하는 것이 바람직하다. 전체 대기시간은 오디오 데이터의 왕복에서 모든 대기시간의 합에 기반하여 결정될 수 있다. 예를 들어, 전체 대기시간은 왕복 대기시간, 인코딩 대기시간, 디코딩 대기시간 및 렌더링 대기시간을 기반으로 할 수 있다.

[0100] 이 대기시간을 숨기는 것은 (예를 들어, S920단계 및 S1020 단계를 참조하여 상술한 바와 같이) 렌더링을 위해 수신기로부터 송신기/서버로 로컬 포즈를 전송하고 송신기/서버가 렌더링된 콘텐츠 프레임 각각에 대해 사용된 포즈를 되돌려 보내도록 함으로써 달성될 수 있다. 송신기/송신기는 이전에 수신된 위치를 고려하는 것을 포함하여, 콘텐츠가 송신기에 의해 렌더링되고 수신기에서 수신되는 시간 사이에 도입된 추가 대기시간을 보상하기 위하여 사용자의 움직임을 예측할 수 있다.

[0101] 그런 다음, 수신기는 (예를 들어, 단계 S970 및 단계 S1070을 참조하여 상술한 바와 같이) 송신기 측에서 콘텐츠를 렌더링하는 데 사용된 포즈와 수신기(R)의 로컬 포즈(예를 들어, 현재 또는 실제 포즈) 사이의 델타가 주어지면, 서버로부터 수신한 사전 렌더링된 오디오를 외삽할 수 있다.

[0102] 이 외삽은 렌더링된 콘텐츠의 유연성을 기반으로 여러 방법으로 구현될 수 있다. 일 예에서, 콘텐츠가 사전 렌더링된 앰비소닉 B-포맷이고 모션이 3 자유도 모션인 경우, 외삽은 재생 전에 FOA 또는 B-포맷 콘텐츠의 클라이언트 측 로컬 회전에 기반할 수 있다. 다른 예에서, 사전 렌더링된 양이 콘텐츠의 경우, 외삽은 블라인드 업믹싱(부록 A 참조) 또는 양이 스트림에 메타데이터 추가(부록 B 참조)를 통해 달성할 수 있다. 다른 예에서, 사전 렌더링된 채널 기반 콘텐츠에 대해, 저 대기시간 블라인드 업믹서가 수신단에 적용될 수 있다.

[0103] 렌더링 및 인코딩이 송신기 측에서 긴밀하게 통합되는 경우, 메타데이터 인코딩, 예를 들어, 현재 렌더링 위치 P, ▽E(P)에 기반하여 에너지 기울기 또는 상이한 부대역의 방향/거리를 추가함으로써 사전 렌더링된 콘텐츠의 유연성을 증가시킬 수 있다.

[0104] 렌더링될 원본 콘텐츠가 객체 기반이면, 원하는 위치 주변에서 다수의 렌더링을 계산하고 레벨 기울기를 인코딩할 수 있다. 이 레벨 기울기 G는 일반적으로 3D 벡터(세 축 x, y, z 각각에 대해 하나의 값)로 구성된다. 수신기는 사전 렌더링된 위치 P와 현재 수신기 위치 P' 사이의 차이에 기반하여 수신된 신호의 부대역 에너지 E(P)를 $E(P') = E(P) \cdot (P' - P) \cdot \nabla E(P)$와 같이 간단히 조정할 수 있다.

[0105] 이 추가 정보는, 예를 들어 (거리 정보를 사용하여) 시차 효과를 계산하거나 (레벨 기울기 정보를 사용하여) 렌더링 레벨을 조정하는 것과 같이 사전 렌더링된 스트림(즉 사전 렌더링된 미디어 콘텐츠)을 추가로 외삽하기 위해 수신기에서 사용될 수 있다.

[0106] 일 예에서, 수신기가 계산 능력의 관점에서 제약을 받는다면, 업믹싱이 송신기 측에서 인코딩 중에 수행될 수 있다. 예를 들어, B-포맷 또는 채널은 객체로 변환될 수 있다. 이는 인코딩 경로 대기시간을 증가시킬 수 있지만, 결과적인 콘텐츠가 더 유연해지고 수신기 측에서 외삽될 수 있다.

[0107] 사용자 동작(예를 들어 버튼 트리거)이 게임 플레이에 영향을 줄 수 있는 게임용으로는, 전체 시스템 대기시간이 여전히 20 msec 미만일 필요가 있으며, 이는 복잡한 업믹싱 작업의 실행을 방지할 수 있다. 결과적으로, B-포맷과 같은 유연한 포맷은 수신단에서 저 지연시간으로 렌더링 및 회전될 수 있기 때문에 저 지연시간 무손실 또는 손실 코덱을 사용하는 렌더링/전송에 가장 적합한 후보가 될 수 있다.

[0108] 다양한 오디오 코덱은 상술한 데이터 전송 모드를 통합할 수 있다. 코덱은 다음에 대해 적응될 수 있다: (i) 무손실 부호화된 (제로 대기시간 부호화된) 스테레오 오디오 데이터 또는 저 대기 시간 손실 데이터를 전송할 가능성; (ii) "통상적인" 렌더링(예를 들어 디바이스 내에서의 양이화)이 해제되어야 하는(예를 들어 비트스트림 신택스 필드. Dolby AC-4 및 MPEG-H Part 3, 3D Audio 양자가 예컨대 Dolby AC-4에서 b_pre_virtualized와 같은 이러한 비트필드를 이미 포함하는) 경우, 콘텐츠가 이미 사전 렌더링되었음을 신호하는 수단; 및 (iii) 필요한 경우, HRTFs 및 BRIRs를 전송하는 수단.

[0109] 따라서, 본 개시의 문맥에서, 송신기는 또한 사전 렌더링된 오디오 콘텐츠를 제공한다는 표시(예를 들어, 플래그, 비트 필드, 신택스 필드/엘리먼트, 매개변수)를 수신기에 제공할 수 있다. 그러한 표시가 수신기에 의해 수신되면, 수신기는 오디오 콘텐츠의 임의의 (수신기 측) 렌더링을 포기할 수 있다. 예를 들어, 양이 사전 렌더링된 오디오 콘텐츠의 경우, 수신기는 더 이상 렌더링하지 않고, 송신기로부터 수신한 사전 렌더링된 오디오 콘텐츠를 재현용 헤드폰(의 스피커)으로 직접 라우팅할 수 있다. 이러한 표시는 비트스트림 내에서 수신기에 신호되는 매개변수 directHeadphone의 형태일 수 있다. directHeadphone 매개변수는 양이 출력이 렌더링되면, 채널(유형)의 해당하는 신호 그룹이 헤드폰 출력으로 직접 이동하도록 정의할 수 있다. 신호는 왼쪽 및 오른쪽 헤드폰 채널로 라우팅될 수 있다.

[0110] 이 매개변수에 대한 신택스의 가능한 예가 표 4에 재현된다.

표 4

[0111]

신택스	비트 수	니모닉
prodMetadataConfig() { /* high resolution object distance */ hasObjectDistance;	1	bslbf
/* direct to headphone */ for (gp = 0; gp < numChannelGroups; gp++) { directHeadphone [gp];	1	bslbf
} }		

[0112] 시맨틱은 MPEG-H 3D 오디오 (ISO/IEC 23008-3) 및/또는 MPEG 표준의 미래 버전에 따라 정의될 수 있다.

[0113] 본 개시의 다른 예시적인 실시예는 아래에 기재된 열거된 예시적 실시예(Enumerated Example Embodiments,

EEE)에 요약되어 있다.

[0114] 제1 EEE는 제1 장치에 의한 재현을 위하여 미디어 콘텐츠를 처리하는 방법에 관한 것으로서, 방법은: 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하는 것, 포즈 정보를 미디어 콘텐츠를 제공하는 제2 장치로 전송하는 것, 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하는 것, 및 렌더링된 미디어 콘텐츠를 재현을 위하여 제1 장치로 전송하는 것을 포함한다.

[0115] 제2 EEE는 제1 EEE의 방법에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며; 및/또는 미디어 콘텐츠는 비디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 비디오 콘텐츠를 포함한다.

[0116] 제3 EEE는 제1 EEE의 방법에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 방법은 렌더링된 오디오 콘텐츠의 가청 표현을 생성하는 것을 더 포함한다.

[0117] 제4 EEE는 제2 EEE의 방법에 관한 것으로서, 오디오 콘텐츠는 일차 앰비소닉(First Order Ambisonics: FOA) 기반, 고차 앰비소닉(Higher Order Ambisonics: HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0118] 제5 EEE는 제2 EEE의 방법에 관한 것으로서, 렌더링된 오디오 콘텐츠는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0119] 제6 EEE는 제1 EEE 내지 제5 EEE 중 어느 하나의 방법에 관한 것으로서, 렌더링하는 것은: 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보를 획득하는 것, 및 렌더링된 미디어 콘텐츠를 획득하기 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하는 것을 수반한다.

[0120] 제7 EEE는 제6 EEE의 방법에 관한 것으로서, 렌더링된 미디어 콘텐츠와 함께 예측 포즈 정보를 제1 장치로 전송하는 것을 더 포함한다.

[0121] 제8 EEE는 제7 EEE의 방법에 관한 것으로서, 예측 포즈 정보를 실제 포즈 정보와 비교하는 것, 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 더 포함한다.

[0122] 제9 EEE는 제8 EEE의 방법에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측되며, 실제 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 실제로 처리되는 타이밍에 획득된 포즈 정보이다.

[0123] 제10 EEE는 제1 EEE 내지 제9 EEE중 어느 하나의 방법에 관한 것으로서, 렌더링된 미디어 콘텐츠는 비압축 형태로 상기 제1 장치로 전송된다.

[0124] 제11 EEE는 제1 EEE 내지 제10 EEE중 어느 하나의 방법에 관한 것으로서, 제1 장치로의 전송 전에 렌더링된 미디어 콘텐츠를 인코딩하는 것; 및 제1 장치에서의 수신 후에 인코딩된 렌더링된 미디어 콘텐츠를 디코딩하는 것을 더 포함한다.

[0125] 제12 EEE는 제9 EEE 또는 제9 EEE의 특징을 포함하는 임의의 EEE 의 방법에 관한 것으로서, 렌더링된 오디오 콘텐츠가 재현을 위하여 상기 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정은 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 상기 제1 장치로 전송하는 데 필요한 시간의 추정을 포함한다.

[0126] 제13 EEE는 제6 EEE 또는 제6 EEE의 특징을 포함하는 임의의 EEE 의 방법에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 상기 제1 장치로 전송하는 데 필요한 시간의 추정에 추가로 기반하여 획득된다.

[0127] 제14 EEE는 제1 EEE 내지 제13 EEE 중 어느 하나의 방법에 관한 것으로서, 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하는 것, 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 더 포함한다.

[0128] 제15 EEE는 제1 EEE 내지 제14 EEE 중 어느 하나의 방법에 관한 것으로서, 제2 장치에서, 포즈 정보의 변화에 응답하여 렌더링된 미디어 콘텐츠가 어떻게 변화하는지를 나타내는 기울기 정보를 결정하는 것, 렌더링된 미디

어 콘텐츠와 함께 기울기 정보를 제1 장치로 전송하는 것, 제1 장치에서, 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하는 것, 및 기울기 정보 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하는 것을 더 포함한다.

[0129] 제16 EEE는 제1 EEE 내지 제15 EEE 중 어느 하나의 방법에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 방법은 제1 장치가 위치한 환경의 음향 특성을 나타내는 환경 정보를 제2 장치로 전송하는 것을 더 포함하고, 미디어 콘텐츠의 렌더링은 환경 정보에 추가로 기반한다.

[0130] 제17 EEE는 제1 EEE 내지 제16 EEE 중 어느 하나의 방법에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 방법은 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보를 제2 장치로 전송하는 것을 더 포함하고, 미디어 콘텐츠의 렌더링은 형태 정보에 추가로 기반한다.

[0131] 제18 EEE는 미디어 콘텐츠를 재현하는 제1 장치 및 미디어 콘텐츠를 저장하는 제2 장치를 포함하는 시스템에 관한 것으로서, 제1 장치는 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하고, 포즈 정보를 제2 장치로 전송하도록 적응되고, 제2 장치는 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하고, 렌더링된 미디어 콘텐츠를 재현하기 위하여 제1 장치로 전송하도록 적응된다.

[0132] 제19 EEE는 제18 EEE의 시스템에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 및/또는 미디어 콘텐츠는 비디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 비디오 콘텐츠를 포함한다.

[0133] 제20 EEE는 제18 EEE의 시스템에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 렌더링된 오디오 콘텐츠의 가청 표현을 생성하도록 더 적응된다.

[0134] 제21 EEE는 제19 EEE의 시스템에 관한 것으로서, 오디오 콘텐츠는 일차 앰비소닉(FOA) 기반, 고차 앰비소닉(HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0135] 제22 EEE는 제19 EEE 내지 제21 EEE 중 어느 하나의 시스템에 관한 것으로서, 렌더링된 오디오 콘텐츠는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0136] 제23 EEE는 제18 EEE 내지 제22 EEE 중 어느 하나의 시스템에 관한 것으로서, 제2 장치는 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보를 획득하고, 렌더링된 미디어 콘텐츠를 획득하기 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하도록 더 적응된다.

[0137] 제24 EEE는 제23 EEE의 시스템에 관한 것으로서, 제2 장치는 렌더링된 미디어 콘텐츠와 함께 예측 포즈 정보를 제1 장치로 전송하도록 더 적응된다.

[0138] 제25 EEE는 제24 EEE의 시스템에 관한 것으로서, 제1 장치는 예측 포즈 정보를 실제 포즈 정보와 비교하고, 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0139] 제26 EEE는 제25 EEE의 시스템에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측되며, 실제 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 실제로 처리되는 타이밍에 획득된 포즈 정보이다.

[0140] 제27 EEE는 제18 EEE 내지 제26 EEE 중 어느 하나의 시스템에 관한 것으로서, 렌더링된 미디어 콘텐츠는 비압축 형태로 상기 제1 장치로 전송된다.

[0141] 제28 EEE는 제18 EEE 내지 제27 EEE 중 어느 하나의 시스템에 관한 것으로서, 제2 장치는 제1 장치로의 전송 전에 렌더링된 미디어 콘텐츠를 인코딩하도록 더 적응되고, 제1 장치는 제1 장치에서의 수신 후에 상기 인코딩된 렌더링된 미디어 콘텐츠를 디코딩하도록 더 적응된다.

[0142] 제29 EEE는 제26 EEE 또는 제26 EEE의 특징을 포함하는 임의의 EEE 의 시스템에 관한 것으로서, 렌더링된 오디오 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정은 렌더링된 오디오 콘텐츠

의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정을 포함한다.

[0143] 제30 EEE는 제23 EEE 또는 제23 EEE의 특징을 포함하는 임의의 EEE 의 시스템에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정에 추가로 기반하여 획득된다.

[0144] 제31 EEE는 제18 EEE 내지 제30 EEE 중 어느 하나의 시스템에 관한 것으로서, 제1 장치는 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하고, 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0145] 제32 EEE는 제18 EEE 내지 제31 EEE 중 어느 하나의 시스템에 관한 것으로서, 제2 장치는 포즈 정보의 변화에 응답하여 렌더링된 미디어 콘텐츠가 어떻게 변화하는지를 나타내는 기울기 정보를 결정하고, 렌더링된 미디어 콘텐츠와 함께 기울기 정보를 제1 장치로 전송하도록 더 적응되고, 제1 장치는 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하고, 기울기 정보 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0146] 제33 EEE는 제18 EEE 내지 제32 EEE 중 어느 하나의 시스템에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 제1 장치가 위치한 환경의 음향 특성을 나타내는 환경 정보를 제2 장치로 전송하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 환경 정보에 추가로 기반한다.

[0147] 제34 EEE는 제18 EEE 내지 제33 EEE 중 어느 하나의 시스템에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보를 제2 장치로 전송하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 형태 정보에 추가로 기반한다.

[0148] 제35 EEE는 제1 장치에 의한 재현을 위하여 미디어 콘텐츠를 제공하는 제2 장치에 관한 것으로서, 제2 장치는 제1 장치의 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 수신하고, 렌더링된 미디어 콘텐츠를 획득하기 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하고, 렌더링된 미디어 콘텐츠를 재현을 위하여 제1 장치로 전송하도록 적응된다.

[0149] 제36 EEE는 제35 EEE의 제2 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 및/또는 미디어 콘텐츠는 비디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 비디오 콘텐츠를 포함한다.

[0150] 제37 EEE는 제36 EEE의 제2 장치에 관한 것으로서, 오디오 콘텐츠는 일차 앰비소닉(FOA) 기반, 고차 앰비소닉(HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0151] 제38 EEE는 제36 EEE의 제2 장치에 관한 것으로서, 렌더링된 오디오 콘텐츠는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0152] 제39 EEE는 제35 EEE 내지 제38 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보를 획득하고, 렌더링된 미디어 콘텐츠를 획득하기 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하도록 더 적응된다.

[0153] 제40 EEE는 제39 EEE의 제2 장치에 관한 것으로서, 렌더링된 미디어 콘텐츠와 함께 예측 포즈 정보를 제1 장치로 전송하도록 더 적응된다.

[0154] 제41 EEE는 제39 EEE 또는 제40 EEE의 제2 장치에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측된다.

[0155] 제42 EEE는 제35 EEE 내지 제41 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 렌더링된 미디어 콘텐츠는 비압축 형태로 제1 장치로 전송된다.

[0156] 제43 EEE는 제35 EEE 내지 제42 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 제1 장치로의 전송 전에 렌더링된 미디어 콘텐츠를 인코딩하도록 더 적응된다.

[0157] 제44 EEE는 제41 EEE 또는 제41 EEE의 특징을 포함하는 임의의 EEE 의 제2 장치에 관한 것으로서, 렌더링된 오디오 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정은 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정을 포함한다.

[0158] 제45 EEE는 제39 EEE 또는 제39 EEE의 특징을 포함하는 임의의 EEE 의 제2 장치에 관한 것으로서, 예측 포즈 정보는 렌더링된 미디어 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정에 추가로 기반하여 획득된다.

[0159] 제46 EEE는 제35 EEE 내지 제45 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 포즈 정보의 변화에 응답하여 렌더링된 미디어 콘텐츠가 어떻게 변화하는지를 나타내는 기울기 정보를 결정하고, 렌더링된 미디어 콘텐츠와 함께 기울기 정보를 제1 장치로 전송하도록 더 적응된다.

[0160] 제47 EEE는 제35 EEE 내지 제46 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제2 장치는 제1 장치가 위치한 환경의 음향 특성을 나타내는 환경 정보를 제1 장치로부터 수신하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 환경 정보에 추가로 기반한다.

[0161] 제48 EEE는 제35 EEE 내지 제47 EEE 중 어느 하나의 제2 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제2 장치는 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보를 제1 장치로부터 수신하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 형태 정보에 추가로 기반한다.

[0162] 제49 EEE는 제2 장치에 의해 제공되는 미디어 콘텐츠를 재현하는 제1 장치에 관한 것으로서, 제1 장치는: 제1 장치의 사용자의 위치 및/또는 방향을 나타내는 포즈 정보를 획득하고, 포즈 정보를 제2 장치로 전송하고, 제2 장치로부터 렌더링된 미디어 콘텐츠를 수신-렌더링된 미디어 콘텐츠는 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링하여 획득됨-하고, 렌더링된 미디어 콘텐츠를 재현하도록 적응된다.

[0163] 제50 EEE는 제49 EEE의 제1 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 및/또는 미디어 콘텐츠는 비디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 비디오 콘텐츠를 포함한다.

[0164] 제51 EEE는 제49 EEE의 제1 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 렌더링된 오디오 콘텐츠의 가청 표현을 생성하도록 더 적응된다.

[0165] 제52 EEE는 제50 EEE 또는 제51 EEE의 제1 장치에 관한 것으로서, 오디오 콘텐츠는 일차 앰비소닉(FOA) 기반, 고차 앰비소닉(HOA) 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 FOA 기반, HOA 기반, 객체 기반 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0166] 제53 EEE는 제50 EEE 내지 제52 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 렌더링된 오디오 콘텐츠는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠 중 하나이거나, 또는 양이 오디오 콘텐츠, FOA 오디오 콘텐츠, HOA 오디오 콘텐츠 또는 채널 기반 오디오 콘텐츠의 둘 이상의 조합이다.

[0167] 제54 EEE는 제49 EEE 내지 제53 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 제2 장치로부터 렌더링된 미디어 콘텐츠와 함께 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 수신하고, 미디어를 렌더링하는 데 사용된 포즈 정보를 실제 포즈 정보와 비교하고, 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0168] 제55 EEE는 제54 EEE의 제1 장치에 관한 것으로서, 실제 포즈 정보는 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리되는 타이밍에 획득된 포즈 정보이다.

[0169] 제56 EEE는 제49 EEE 내지 제55 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보를 획득하고, 예측 포즈 정보를 제2 장치로 전송하도록 더 적응된다.

[0170] 제57 EEE는 제56 EEE의 제1 장치에 관한 것으로서, 예측 포즈 정보는 상기 렌더링된 미디어 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정에 대해 예측된다.

[0171] 제58 EEE는 제49 EEE 내지 제57 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 렌더링된 미디어 콘텐츠는 비압축 형태로 제2 장치로부터 수신된다.

[0172] 제59 EEE는 제49 EEE 내지 제58 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 제1 장치는 인코딩된 렌더링된 미디어 콘텐츠를 디코딩하도록 더 적응된다.

[0173] 제60 EEE는 제57 EEE 또는 제57 EEE의 특징을 포함하는 임의의 EEE 의 제2 장치에 관한 것으로서, 렌더링된 오디오 콘텐츠가 재현을 위하여 제1 장치에 의해 처리될 것으로 예상되는 타이밍의 추정은 렌더링된 오디오 콘텐츠의 인코딩 및 디코딩에 필요한 시간의 추정 및/또는 렌더링된 미디어 콘텐츠를 제1 장치로 전송하는 데 필요한 시간의 추정을 포함한다.

[0174] 제61 EEE는 제49 EEE 내지 제60 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하고, 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0175] 제62 EEE는 제49 EEE 내지 제61 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 제2 장치로부터 렌더링된 미디어 콘텐츠와 함께 포즈 정보의 변화에 응답하여 렌더링된 미디어 콘텐츠가 어떻게 변화하는지를 나타내는 기울기 정보를 수신하고, 미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교하고, 기울기 정보 및 비교의 결과에 기반하여 렌더링된 미디어 콘텐츠를 업데이트하도록 더 적응된다.

[0176] 제63 EEE는 제49 EEE 내지 제62 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 제1 장치가 위치한 환경의 음향 특성을 나타내는 환경 정보를 제2 장치로 전송하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 환경 정보에 추가로 기반한다.

[0177] 제64 EEE는 제49 EEE 내지 제63 EEE 중 어느 하나의 제1 장치에 관한 것으로서, 미디어 콘텐츠는 오디오 콘텐츠를 포함하고 렌더링된 미디어 콘텐츠는 렌더링된 오디오 콘텐츠를 포함하며, 제1 장치는 사용자 또는 사용자의 일부의 형태를 나타내는 형태 정보를 제2 장치로 전송하도록 더 적응되고, 미디어 콘텐츠의 렌더링은 형태 정보에 추가로 기반한다.

[0178] 제65 EEE는 오디오 콘텐츠를 렌더링하는 방법에 관한 것이다. 방법은 송신기(S) 장치에 의하여, 사용자 위치 및/또는 방향 데이터를 수신하는 것 및, 통상적으로 객체-5 기반 또는 HOA 표현으로부터 유도되는, 대응하는 사전 렌더링된 콘텐츠를 송신하는 것을 포함한다.

[0179] 제66 EEE는 제65 EEE의 방법에 관한 것으로서, 송신기에 의해 생성되는 사전 렌더링된 신호는 양이, FOA/B-포맷, HOA 또는 임의 유형의 채널 기반 렌더링이다.

[0180] 제67 EEE는 제65 EEE 또는 제66 EEE 의 방법에 관한 것으로서, 비압축 사전 렌더링된 콘텐츠를 전송하는 것을 더 포함한다.

[0181] 제68 EEE는 제65 EEE 또는 제66 EEE 의 방법에 관한 것으로서, 사전 렌더링된 콘텐츠를 인코딩하는 것 및 인코딩된 사전 렌더링된 콘텐츠를 전송하는 것을 더 포함한다.

[0182] 제69 EEE는 제65 EEE 내지 제68 EEE 중 어느 하나의 방법에 관한 것으로서, 수신기에 의하여 사전 렌더링된 콘텐츠를 수신하는 것을 더 포함한다.

[0183] 제70 EEE는 제65 EEE 내지 제69 EEE 중 어느 하나의 방법에 관한 것으로서, 수신기에 의해 사전 렌더링된, 사전 인코딩된 양이화된 콘텐츠를 디코딩하는 것을 더 포함한다.

[0184] 제71 EEE는 제65 EEE 내지 제70 EEE 중 어느 하나의 방법에 관한 것으로서, 사용자 위치 및/또는 방향 데이터는 월드 스페이서(world space) 내의 사용자의 위치 및 방향을 나타내는 로컬 포즈를 포함한다.

[0185] 제72 EEE는 제65 EEE 내지 제71 EEE 중 어느 하나의 방법에 관한 것으로서, 사용자 위치 데이터는 수신기로부터 송신기로 전송된다.

[0186] 제73 EEE는 제65 EEE 내지 제72 EEE 중 어느 하나의 방법에 관한 것으로서, 사전 렌더링된 양이화된 콘텐츠를 위해 사용된 사용자 위치 데이터를 수신기로 다시 전송하는 것을 더 포함한다.

[0187] 제74 EEE는 제65 EEE 내지 제73 EEE 중 어느 하나의 방법에 관한 것으로서, 수신된 사용자 위치 데이터 및 로컬 위치 데이터에 기반하여 사전 렌더링된 콘텐츠를 외삽하여 업데이트된 콘텐츠를 결정하는 것을 더 포함한다.

[0188] 제75 EEE는 제65 EEE 내지 제74 EEE 중 어느 하나의 방법에 관한 것으로서, 개별화된 양이 처리를 위하여 사용자에 대한 형태 데이터(예를 들어 머리 크기, 머리 형상)를 전송하는 것을 더 포함한다.

[0189] 제76 EEE는 제65 EEE 내지 제75 EEE 중 어느 하나의 방법에 관한 것으로서, BRIR 함수 및/또는 룸 특성에 대한 데이터를 전송하는 것을 더 포함한다.

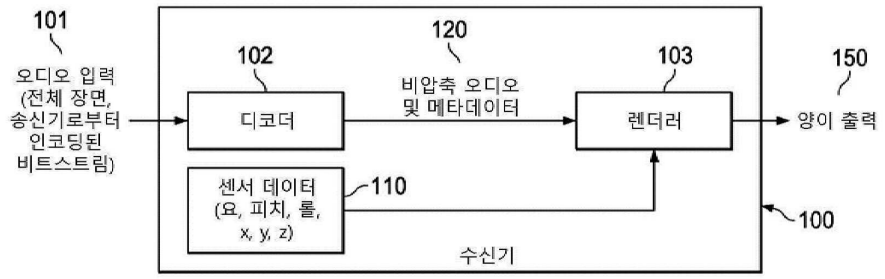
[0190] 제77 EEE는 제65 EEE 내지 제76 EEE 중 어느 하나의 방법에 관한 것으로서, 콘텐츠가 청취자를 알 수 없는 방식(예를 들어, HRTFs를 포함하지 않음)으로 전송된다는 결정에 기반하여, 수신기 측에서 양이 렌더링 및 개별화를 수행하는 것을 더 포함한다.

[0191] 제78 EEE는 제65 EEE 내지 제77 EEE 중 어느 하나의 방법에 관한 것으로서, 시점 t1에서 사용자 위치 및/또는 방향 데이터 P(t0)를 제공하는 것을 더 포함한다.

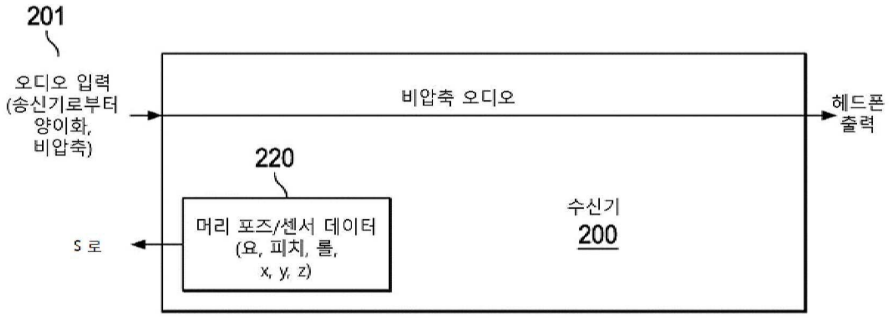
[0192] 제79 EEE는 제67 EEE의 방법에 관한 것으로서, 비압축 사전 렌더링된 콘텐츠는 양이화된 비압축 사전 렌더링된 콘텐츠이다.

도면

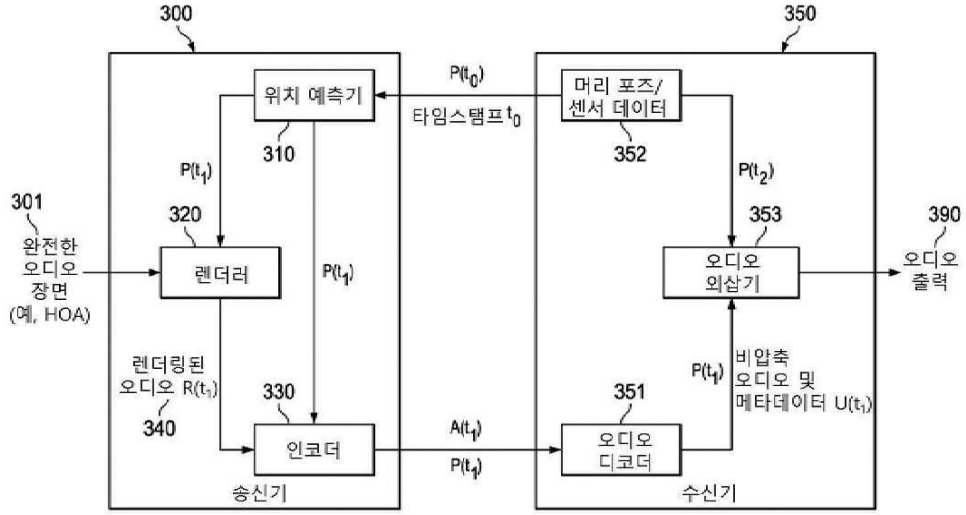
도면1



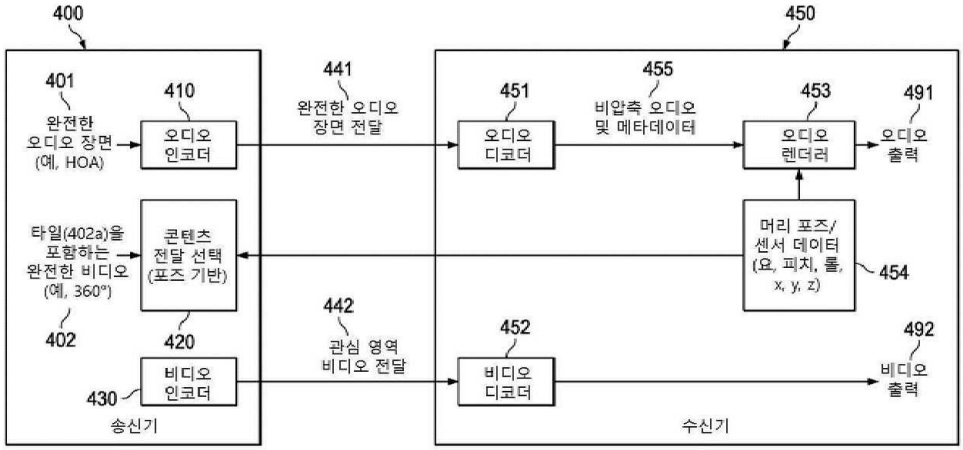
도면2



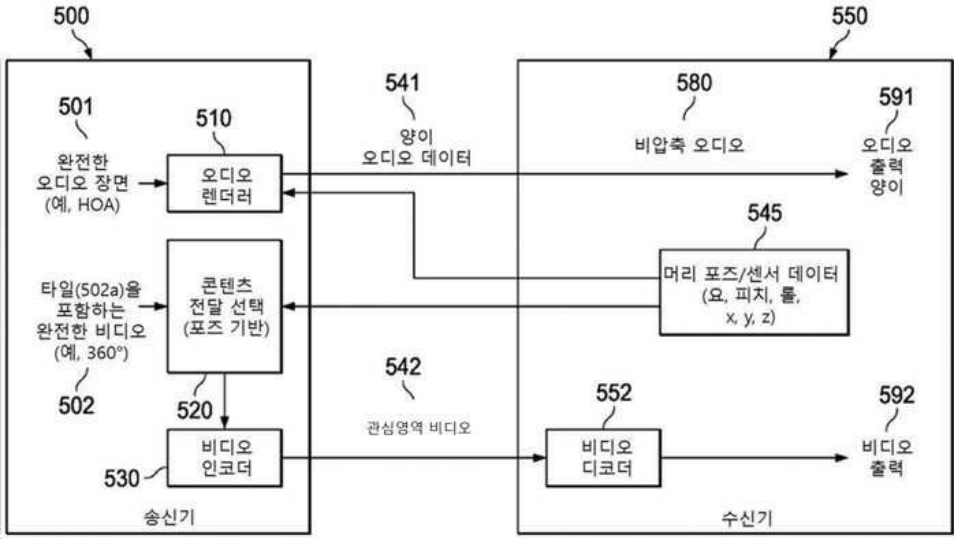
도면3


300
350
위치 예측기
310
$P(t_0)$
타임스탬프 t_0
머리 포즈/
센서 데이터
352
$P(t_1)$
$P(t_2)$
301
완전한
오디오
장면
(예, HOA)
320
렌더러
$P(t_1)$
353
오디오
외삽기
390
오디오
출력
렌더링된
오디오 $R(t_1)$
340
330
인코더
$A(t_1)$
$P(t_1)$
351
오디오
디코더
$P(t_1)$
비압축
오디오 및
메타데이터 $U(t_1)$
송신기
수신기

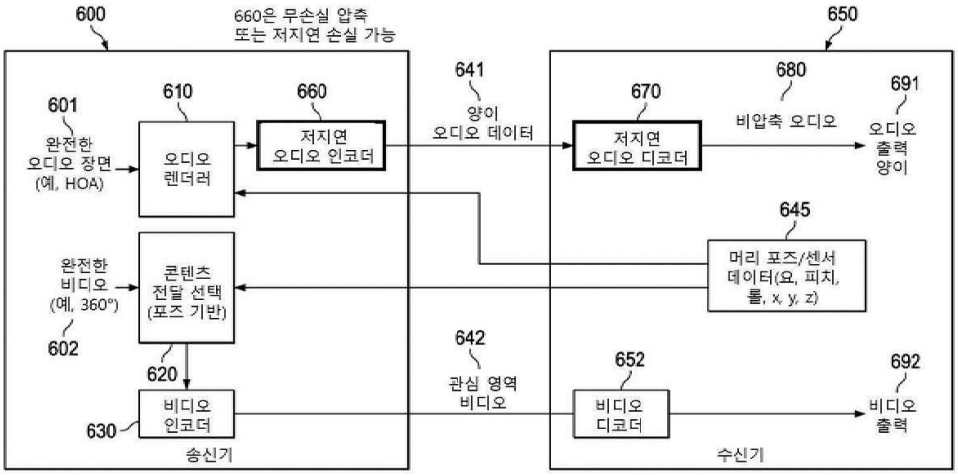
도면4


400
450
401
완전한
오디오 장면
(예, HOA)
410
오디오
인코더
441
완전한 오디오
장면 전달
451
오디오
디코더
455
비압축 오디오
및 메타데이터
453
오디오
렌더러
491
오디오
출력
타일(402a)을
포함하는
완전한 비디오
(예, 360°)
402
콘텐츠
전달 선택
(포즈 기반)
420
머리 포즈/
센서 데이터
(요, 피치, 롤,
x, y, z)
454
442
관심 영역
비디오 전달
452
492
비디오
출력
430
비디오
인코더
비디오
디코더
송신기
수신기

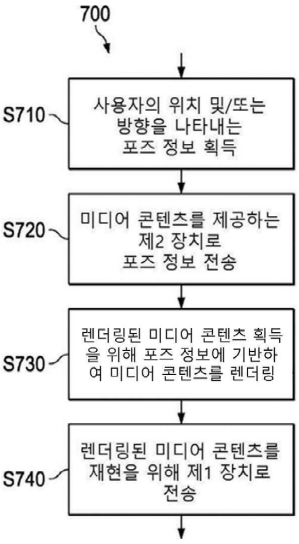
도면5


500
550
501
완전한
오디오 장면
(예, HOA)
510
오디오
렌더러
541
양이
오디오 데이터
580
비압축 오디오
591
오디오
출력
양이
545
머리 포즈/센서 데이터
(요, 피치, 롤,
x, y, z)
타일(502a)을
포함하는
완전한 비디오
(예, 360°)
502
콘텐츠
전달 선택
(포즈 기반)
520
542
관심영역 비디오
552
592
비디오
출력
530
비디오
인코더
비디오
디코더
송신기
수신기

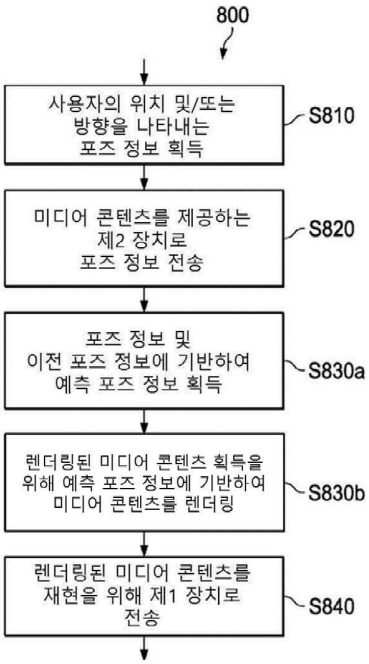
도면6


600
660은 무손실 압축
또는 저지연 손실 가능
650
641
양이
오디오 데이터
680
비압축 오디오
691
오디오
출력
양이
601
완전한
오디오 장면
(예, HOA)
610
오디오
렌더러
660
저지연
오디오 인코더
670
저지연
오디오 디코더
645
머리 포즈/센서
데이터(요, 피치,
롤, x, y, z)
완전한
비디오
(예, 360°)
602
콘텐츠
전달 선택
(포즈 기반)
642
관심 영역
비디오
652
620
비디오
인코더
630
비디오
디코더
692
비디오
출력
송신기
수신기

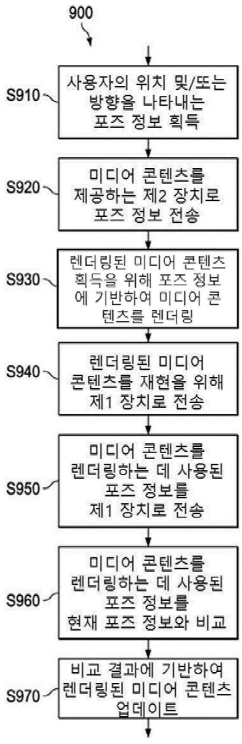
도면7


700
S710
사용자의 위치 및/또는
방향을 나타내는
포즈 정보 획득
S720
미디어 콘텐츠를 제공하는
제2 장치로
포즈 정보 전송
S730
렌더링된 미디어 콘텐츠 획득
을 위해 포즈 정보에 기반하
여 미디어 콘텐츠를 렌더링
S740
렌더링된 미디어 콘텐츠를
재현을 위해 제1 장치로
전송

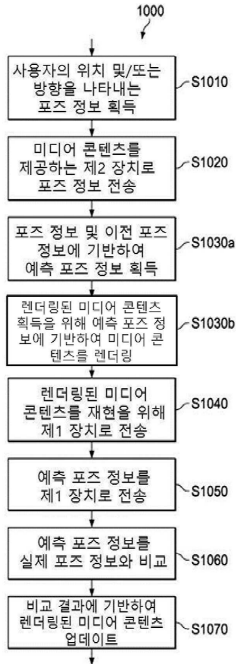
도면8


800
사용자의 위치 및/또는 방향을 나타내는 포즈 정보 획득
S810
미디어 콘텐츠를 제공하는 제2 장치로 포즈 정보 전송
S820
포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보 획득
S830a
렌더링된 미디어 콘텐츠 획득을 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링
S830b
렌더링된 미디어 콘텐츠를 재현을 위해 제1 장치로 전송
S840

도면9


900
S910
사용자의 위치 및/또는 방향을 나타내는 포즈 정보 획득
S920
미디어 콘텐츠를 제공하는 제2 장치로 포즈 정보 전송
S930
렌더링된 미디어 콘텐츠 획득을 위해 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링
S940
렌더링된 미디어 콘텐츠를 재현을 위해 제1 장치로 전송
S950
미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 제1 장치로 전송
S960
미디어 콘텐츠를 렌더링하는 데 사용된 포즈 정보를 현재 포즈 정보와 비교
S970
비교 결과에 기반하여 렌더링된 미디어 콘텐츠 업데이트

도면10


1000
사용자의 위치 및/또는 방향을 나타내는 포즈 정보 획득
S1010
미디어 콘텐츠를 제공하는 제2 장치로 포즈 정보 전송
S1020
포즈 정보 및 이전 포즈 정보에 기반하여 예측 포즈 정보 획득
S1030a
렌더링된 미디어 콘텐츠 획득을 위해 예측 포즈 정보에 기반하여 미디어 콘텐츠를 렌더링
S1030b
렌더링된 미디어 콘텐츠를 재현을 위해 제1 장치로 전송
S1040
예측 포즈 정보를 제1 장치로 전송
S1050
예측 포즈 정보를 실제 포즈 정보와 비교
S1060
비교 결과에 기반하여 렌더링된 미디어 콘텐츠 업데이트
S1070