



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0078537
(43) 공개일자 2020년07월01일

(51) 국제특허분류(Int. Cl.)
H04N 21/439 (2011.01) H04N 21/218 (2011.01)
H04N 21/2343 (2011.01) H04N 21/6587 (2011.01)
H04N 21/81 (2011.01)
(52) CPC특허분류
H04N 21/439 (2013.01)
H04N 21/21805 (2013.01)
(21) 출원번호 10-2020-7013475
(22) 출원일자(국제) 2018년10월11일
심사청구일자 2020년06월01일
(85) 번역문제출일자 2020년05월11일
(86) 국제출원번호 PCT/EP2018/077770
(87) 국제공개번호 WO 2019/072984
국제공개일자 2019년04월18일
(30) 우선권주장
17196259.0 2017년10월12일
유럽특허청(EPO)(EP)

(71) 출원인
프라운호퍼 게젤샤프트 쭈르 뢰르데룽 데어 안겐
반텐 포르슘 에. 베.
독일 80686 뮌헨 한자슈트라쎄 27 체
(72) 발명자
머타자, 아드리안
라드브루넨가쎄 1 90403 뉘른베르크 독일
후치스, 하랄드
암셀슈트라쎄 5 91341 로텐바흐 독일
(뒷면에 계속)
(74) 대리인
한상수

전체 청구항 수 : 총 47 항

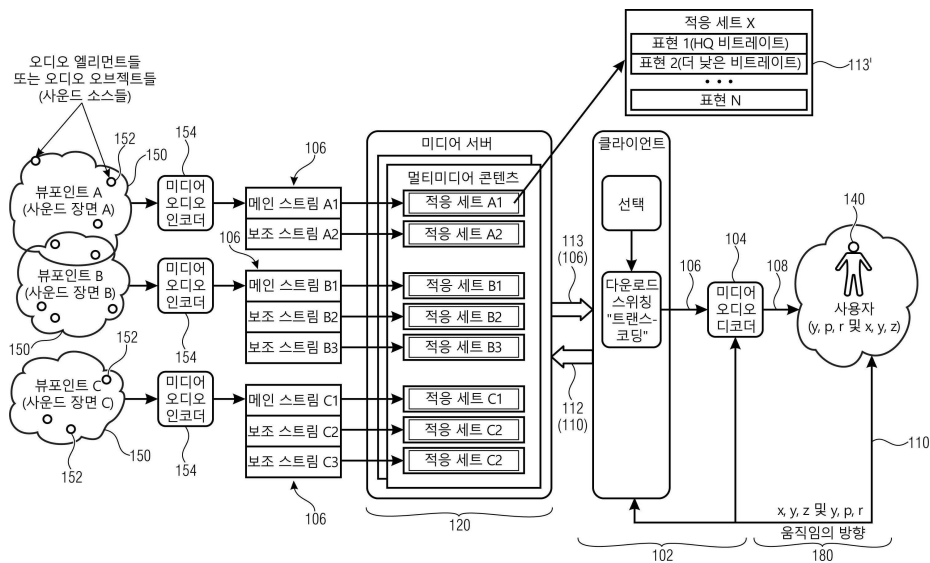
(54) 발명의 명칭 가상 현실 애플리케이션들에 대한 오디오 전달의 최적화

(57) 요약

가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 개시된 기법들, 시스템들, 방법들 및 명령들이 존재한다.

일 예에서, 시스템(102)은 사용자로서의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림(뒷면에 계속)

대표도



들로부터의 비디오 신호들을 디코딩하도록 이루어진 적어도 하나의 미디어 비디오 디코더를 포함한다. 시스템은, 적어도 하나의 오디오 스트림(106)으로부터의 오디오 신호들(108)을 디코딩하도록 이루어진 적어도 하나의 오디오 디코더(104)를 포함한다. 시스템(102)은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 적어도 하나의 오디오 스트림(106) 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 서버(120)에 요청(112)하도록 이루어진다.

(52) CPC특허분류

H04N 21/23439 (2013.01)

H04N 21/6587 (2013.01)

H04N 21/8106 (2013.01)

(72) 발명자

첼한, 베른

루이트폴드스트르. 65 91052 에를랑겐, 독일

프루트스티스, 잔

샌들이테 39a 90768 뤼르트 독일

앙넬리, 마테오

피르크하이머슈트라쎄 11 90408 뉘른베르크 독일

호프만, 인고

캄페슈트라쎄 21 90419 뉘른베르크 독일

명세서

청구범위

청구항 1

미디어 소비 디바이스에서 재생될 비디오 및 오디오 스트림들을 수신하도록 이루어진, 가상 현실(virtual reality; VR), 증강 현실(augmented reality; AR), 혼합 현실(mixed reality; MR), 또는 360도 비디오 환경을 위한 시스템(102)으로서,

상기 시스템(102)은,

사용자로서의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하도록 이루어진 적어도 하나의 미디어 비디오 디코더, 및

적어도 하나의 오디오 스트림(106)으로부터의 오디오 신호들(108)을 디코딩하도록 이루어진 적어도 하나의 오디오 디코더(104)를 포함하며,

상기 시스템(102)은, 상기 사용자의 현재 뷰포트(viewport) 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 상기 적어도 하나의 오디오 스트림(106) 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 서버(120)에 요청(112)하도록 이루어지는, 시스템.

청구항 2

제1항에 있어서,

상기 서버(120)로부터 상기 적어도 하나의 오디오 스트림(106) 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 획득하기 위해, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)를 상기 서버(120)에 제공하도록 이루어지는, 시스템.

청구항 3

제1항 또는 제2항에 있어서,

적어도 하나의 장면이 적어도 하나의 오디오 엘리먼트(152)에 연관되며,

각각의 오디오 엘리먼트는, 상기 장면 내의 상이한 사용자의 위치들 및/또는 뷰포트들 및/또는 머리 배향들 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 대해 상이한 오디오 스트림들이 제공되도록, 상기 오디오 엘리먼트가 가칭적인 시각적 환경 내의 위치 및/또는 영역에 연관되는, 시스템.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 오디오 스트림의 적어도 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트가 상기 장면 내의 현재 사용자의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 대해 재생될지 여부를 판단하도록 이루어지며,

상기 시스템은 현재 사용자의 가상 위치에서 상기 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서,

상기 시스템은, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 상기 오디오 스트림의 적어도 하나의 오디오 엘리먼트(152) 및/또는 하나의 적응 세트가 관련있게 될지 그리고/또는 가청적이게 될지를 예측적으로(predictively) 판단하도록 이루어지고,

상기 시스템은, 상기 장면 내의 예측된 사용자의 움직임 및/또는 상호작용 이전에 특정한 사용자의 가상 위치에서 상기 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림 및/또는 적응 세트를 요청 및/또는 수신하도록 이루어지며,

상기 시스템은, 수신될 경우, 상기 장면 내의 상기 사용자의 움직임 및/또는 상호작용 이후 상기 특정한 사용자의 가상 위치에서 상기 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림을 재생하도록 이루어지는, 시스템.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 장면 내의 상기 사용자의 움직임 및/또는 상호작용 이전에 상기 사용자의 가상 위치에서 더 낮은 비트레이트 및/또는 품질 레벨로 상기 적어도 하나의 오디오 엘리먼트(152)를 요청 및/또는 수신하도록 이루어지며,

상기 시스템은, 상기 장면 내의 상기 사용자의 움직임 및/또는 상호작용 이후 상기 사용자의 가상 위치에서 더 높은 비트레이트 및/또는 품질 레벨로 상기 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서,

상기 적어도 하나의 오디오 엘리먼트(152)는 적어도 하나의 장면에 연관되고, 각각의 오디오 엘리먼트는 상기 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관되며,

상기 시스템은, 상기 사용자로부터 더 멀리있는 오디오 엘리먼트들보다 상기 사용자에게 더 가까운 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질로 스트림들을 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서,

상기 적어도 하나의 오디오 엘리먼트(152)는 적어도 하나의 장면에 연관되고, 상기 적어도 하나의 오디오 엘리먼트는 상기 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관되며,

상기 시스템은, 상기 장면 내의 각각의 사용자의 가상 위치에서의 오디오 엘리먼트들의 관련성 및/또는 가청성 레벨에 기초하여, 상기 오디오 엘리먼트들에 대해 상이한 비트레이트들 및/또는 품질 레벨들로 상이한 스트림들을 요청하도록 이루어지고,

상기 시스템은,

상기 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가청적인 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 오디오 스트림을 요청하고/하거나

상기 현재 사용자의 가상 위치에서 덜 관련있고 덜 가청적인 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 오디오 스트림을

요청하도록 이루어지는, 시스템.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서,

상기 적어도 하나의 오디오 엘리먼트(152)는 장면에 연관되고, 각각의 오디오 엘리먼트는 상기 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관되며,

상기 시스템은, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)를 상기 서버에 주기적으로 전송하여,

제1 위치에 대해, 더 높은 비트레이트 및/또는 품질의 스트림이 상기 서버로부터 제공되고, 그리고

제2 위치에 대해, 더 낮은 비트레이트 및/또는 품질의 스트림이 상기 서버로부터 제공되도록 이루어지고,

상기 제1 위치는 상기 제2 위치보다 상기 적어도 하나의 오디오 엘리먼트(152)에 더 가까운 것인, 시스템.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서,

복수의 장면들(150A, 150B)이 인접한 및/또는 이웃한 환경들과 같은 복수의 시각적 환경들에 대해 정의되어,

제1의 현재 장면에 연관된 제1 스트림들이 제공되고, 제2의 추가적인 장면으로의 사용자의 전환의 경우에는, 상기 제1 장면에 연관된 스트림들 및 상기 제2 장면에 연관된 제2 스트림들 둘 모두가 제공되는, 시스템.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서,

복수의 장면들(150A, 150B)이 제1 시각적 환경 및 제2 시각적 환경에 대해 정의되고, 상기 제1 환경 및 상기 제2 환경은 인접한 및/또는 이웃한 환경들이고,

상기 사용자의 위치 또는 가상 위치가 상기 제1 장면에 연관된 제1 환경에 있는 경우 상기 제1 장면에 연관된 제1 스트림들이 상기 제1 장면의 재생을 위해 상기 서버로부터 제공되고,

상기 사용자의 위치 또는 가상 위치가 상기 제2 장면에 연관된 제2 환경에 있는 경우 상기 제2 장면에 연관된 제2 스트림들이 상기 제2 장면의 재생을 위해 상기 서버로부터 제공되며,

상기 사용자의 위치 또는 가상 위치가 상기 제1 장면과 상기 제2 장면 사이의 전환 위치에 있는 경우 상기 제1 장면에 연관된 제1 스트림들 및 상기 제2 장면에 연관된 제2 스트림들 둘 모두가 제공되는, 시스템.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서,

복수의 장면들(150A, 150B)이 인접한 및/또는 이웃한 환경들인 제1 시각적 환경 및 제2 시각적 환경에 대해 정의되며,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제1 환경에 있는 경우 상기 제1 환경에 연관된 제1 장면(150A)의 재생을 위해 상기 제1 장면에 연관된 제1 스트림들을 요청 및/또는 수신하도록 이루어지고,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제2 환경에 있는 경우 상기 제2 환경에 연관된 제2 장면(150B)의 재생을 위해 상기 제2 장면에 연관된 제2 스트림들을 요청 및/또는 수신하도록 이루어지며,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제1 환경과 상기 제2 환경 사이의 전환 위치(150AB)에 있는 경

우 상기 제1 장면에 연관된 제1 스트림들 및 상기 제2 장면에 연관된 제2 스트림들 둘 모두를 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 13

제10항 내지 제12항 중 어느 한 항에 있어서,

상기 사용자가 상기 제1 장면에 연관된 제1 환경에 있는 경우 상기 제1 장면에 연관된 제1 스트림들은 더 높은 비트레이트 및/또는 품질로 획득되는 반면,

상기 사용자가 상기 제1 장면으로부터 상기 제2 장면으로의 전환 위치의 시작부에 있는 경우 상기 제2 환경에 연관된 상기 제2 장면에 연관된 제2 스트림들은 더 낮은 비트레이트 및/또는 품질로 획득되고,

상기 사용자가 상기 제1 장면으로부터 상기 제2 장면으로의 전환 위치의 끝에 있는 경우, 상기 제1 장면에 연관된 제1 스트림들은 더 낮은 비트레이트 및/또는 품질로 획득되고, 상기 제2 장면에 연관된 제2 스트림들은 더 높은 비트레이트 및/또는 품질로 획득되며,

상기 더 낮은 비트레이트 및/또는 품질은 상기 더 높은 비트레이트 및/또는 품질보다 낮은 것인, 시스템.

청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서,

복수의 장면들(150A, 150B)이 인접한 및/또는 이웃한 환경들과 같은 복수의 환경들에 대해 정의되어,

상기 시스템은, 제1의 현재 환경에 연관된 제1 현재 장면에 연관된 스트림들을 획득하도록 이루어지고,

상기 장면의 경계로부터의 상기 사용자의 위치 또는 가상 위치의 거리가 미리 결정된 임계치 미만인 경우, 상기 시스템은, 제2 장면에 연관된 제2의 인접한 및/또는 이웃한 환경에 연관된 오디오 스트림들을 추가로 획득하는, 시스템.

청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서,

복수의 장면들(150A, 150B)이 복수의 시각적 환경들에 대해 정의되어,

상기 시스템은, 더 높은 비트레이트 및/또는 품질로 현재 장면에 연관된 스트림들을 그리고 더 낮은 비트레이트 및/또는 품질로 제2 장면에 연관된 스트림들을 요청 및/또는 획득하며,

상기 더 낮은 비트레이트 및/또는 품질은 상기 더 높은 비트레이트 및/또는 품질보다 낮은 것인, 시스템.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서,

복수의 N개의 오디오 엘리먼트들이 정의되며,

상기 오디오 엘리먼트들의 위치 또는 영역에 대한 상기 사용자의 거리가 미리 결정된 임계치보다 큰 경우, 상기 N개의 오디오 엘리먼트들은 상기 N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들($M < N$)을 획득하도록 프로세싱되어,

상기 N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 상기 사용자의 거리가 미리 결정된 임계치보다 작은 경우 상기 N개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림이 상기 시스템에 제공되거나, 또는

상기 N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 상기 사용자의 거리가 미리 결정된 임계치보다 큰 경우

상기 M개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림이 상기 시스템에 제공되는, 시스템.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서,

적어도 하나의 시각적 환경 장면은 복수의 N개의 오디오 엘리먼트들($N \geq 2$) 중 적어도 하나에 연관되고, 각각의 오디오 엘리먼트는 시각적 환경 내의 위치 및/또는 영역에 연관되며,

적어도, 상기 복수의 N개의 오디오 엘리먼트들 중 적어도 하나는 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고,

적어도, 상기 복수의 N개의 오디오 엘리먼트들 중 적어도 하나는 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고, 상기 적어도 하나의 표현은, 상기 N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들($M < N$)을 획득하도록 상기 N개의 오디오 엘리먼트들을 프로세싱함으로써 획득되고,

상기 시스템은, 상기 오디오 엘리먼트들이 상기 장면 내의 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가치적인 경우 상기 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 상기 표현을 요청하도록 이루어지며,

상기 시스템은, 상기 오디오 엘리먼트들이 상기 장면 내의 현재 사용자의 가상 위치에서 덜 관련있고 그리고/또는 덜 가치적인 경우 상기 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 상기 표현을 요청하도록 이루어지는, 시스템.

청구항 18

제16항 또는 제17항에 있어서,

상기 사용자의 거리 및/또는 관련성 및/또는 가청성 레벨 및/또는 각도 배향이 미리 결정된 임계치보다 낮은 경우, 상이한 스트림들이 상이한 오디오 엘리먼트들에 대해 획득되는, 시스템.

청구항 19

제1항 내지 제18항 중 어느 한 항에 있어서,

상기 시스템은, 상기 장면 내의 상기 사용자의 배향 및/또는 사용자의 움직임 방향 및/또는 사용자의 상호작용들에 기초하여 스트림들을 요청 및/또는 획득하도록 이루어지는, 시스템.

청구항 20

제1항 내지 제19항 중 어느 한 항에 있어서,

상기 뷰포트는 위치 및/또는 가상 위치 및/또는 움직임 데이터 및/또는 머리에 연관되는, 시스템.

청구항 21

제1항 내지 제20항 중 어느 한 항에 있어서,

상이한 오디오 엘리먼트들이 상이한 뷰포트들에서 제공되며,

상기 시스템은, 하나의 제1 오디오 엘리먼트($S1$)가 뷰포트(160-1) 내에 있는 경우, 상기 뷰포트 내에 있지 않은 제2 오디오 엘리먼트($S2$)보다 더 높은 비트레이트로 상기 제1 오디오 엘리먼트를 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 22

제1항 내지 제21항 중 어느 한 항에 있어서,

제1 오디오 스트림들 및 제2 오디오 스트림들을 요청 및/또는 수신하도록 이루어지며,

상기 제1 오디오 스트림들 내의 제1 오디오 엘리먼트들은 상기 제2 오디오 스트림들 내의 제2 오디오 엘리먼트들보다 더 관련있고 그리고/또는 더 가치적이고,

상기 제1 오디오 스트림들은 상기 제2 오디오 스트림들의 비트레이트 및/또는 품질보다 더 높은 비트레이트 및/또는 품질로 요청 및/또는 수신되는, 시스템.

청구항 23

제1항 내지 제22항 중 어느 한 항에 있어서,

적어도 2개의 시각적 환경 장면들이 정의되며,

적어도 하나의 제1 및 제2 오디오 엘리먼트들은 제1 시각적 환경에 연관된 제1 장면에 연관되고, 적어도 하나의 제3 오디오 엘리먼트는 제2 시각적 환경에 연관된 제2 장면에 연관되고,

상기 시스템은 상기 적어도 하나의 제2 오디오 엘리먼트가 상기 제2 시각적 환경 장면과 부가적으로 연관된다는 것을 설명하는 메타데이터를 획득하도록 이루어지고,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제1 시각적 환경에 있는 경우 상기 적어도 제1 및 제2 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어지고,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제2 시각적 환경 장면에 있는 경우 상기 적어도 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어지며,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제1 시각적 환경 장면과 상기 제2 시각적 환경 장면 사이의 전환에 있는 경우 상기 적어도 제1 및 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어지는, 시스템.

청구항 24

제23항에 있어서,

상기 적어도 하나의 제1 오디오 엘리먼트는 적어도 하나의 오디오 스트림 및/또는 적응 세트에서 제공되고, 상기 적어도 하나의 제2 오디오 엘리먼트는 적어도 하나의 제2 오디오 스트림 및/또는 적응 세트에서 제공되며, 상기 적어도 하나의 제3 오디오 엘리먼트는 적어도 하나의 제3 오디오 스트림 및/또는 적응 세트에서 제공되고,

상기 적어도 제1 시각적 환경 장면은 상기 적어도 하나의 제1 및 제2 오디오 스트림들 및/또는 적응 세트들을 요구하는 완전한 장면으로서 메타데이터에 의해 설명되고, 상기 제2 시각적 환경 장면은 상기 적어도 제1 시각적 환경 장면과 연관된 상기 적어도 하나의 제3 오디오 스트림 및/또는 적응 세트 및 상기 적어도 하나의 제2 오디오 스트림 및/또는 적응 세트들을 요구하는 불완전한 장면으로서 메타데이터에 의해 설명되며,

상기 시스템은, 상기 사용자의 가상 위치가 상기 제2 시각적 환경에 있는 경우, 상기 제1 시각적 환경에 속하는 상기 제2 오디오 스트림 및 상기 제2 시각적 환경과 연관된 상기 제3 오디오 스트림을 새로운 단일 스트림으로 병합하는 것을 허용하도록 상기 메타데이터를 조작하도록 이루어진 메타데이터 프로세서를 포함하는, 시스템.

청구항 25

제1항 내지 제24항 중 어느 한 항에 있어서,

상기 시스템은, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 상기 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 메타데이터를 조작하도록 이루어진 메타데이터 프로세서를 포함하는, 시스템.

청구항 26

제25항에 있어서,

상기 메타데이터 프로세서는, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 상기 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 인에이블링 및/또는 디스에이블링시키도록 이루어지며,

상기 메타데이터 프로세서는, 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 상기 오디오 엘리먼트가 더 이상 재생되지 않을 것이라고 상기 시스템이 판단하는 경우, 상기 적어도 하나의 오디오 디코더 이전에 상기 적어도 하나의 오디오 스트림에서 상기 적어도 하나의 오디오 엘리먼트를 디스에이블링시키도록 이루어지고,

상기 메타데이터 프로세서는, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 상기 오디오 엘리먼트가 재생될 것이라고 상기 시스템이 판단하는 경우, 상기 적어도 하나의 오디오 디코더 이전에 상기 적어도 하나의 오디오 스트림에서 상기 적어도 하나의 오디오 엘리먼트를 인에이블링시키도록 이루어지는, 시스템.

청구항 27

제1항 내지 제26항 중 어느 한 항에 있어서,

상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치에 기초하여 선택된 오디오 엘리먼트들의 디코딩을 디스에이블링시키도록 이루어지는, 시스템.

청구항 28

제1항 내지 제27항 중 어느 한 항에 있어서,

현재 오디오 장면에 연관된 적어도 하나의 제1 오디오 스트림을 이웃한, 인접한 및/또는 미래의 오디오 장면에 연관된 적어도 하나의 스트림에 병합하도록 이루어지는, 시스템.

청구항 29

제1항 내지 제28항 중 어느 한 항에 있어서,

상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 대한 상기 통계 또는 집계된 데이터에 연관된 요청을 상기 서버에 송신하기 위해, 상기 통계 또는 집계된 데이터를 획득 및/또는 수집하도록 이루어지는, 시스템.

청구항 30

제1항 내지 제29항 중 어느 한 항에 있어서,

적어도 하나의 스트림에 연관된 메타데이터에 기초하여 그리고 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 기초하여 상기 적어도 하나의 스트림의 디코딩 및/또는 재생을 비활성화시키도록 이루어지는, 시스템.

청구항 31

제1항 내지 제30항 중 어느 한 항에 있어서,

상기 사용자의 현재 또는 추정된 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여, 선택된 오디오 스트림들의 그룹과 연관된 메타데이터를 조작해서,

재생될 오디오 장면을 구성(compose)하는 오디오 엘리먼트들을 선택하고 그리고/또는 인에이블링시키고 그리고/또는 활성화시키고; 그리고/또는

모든 선택된 오디오 스트림들의 단일 오디오 스트림으로의 병합을 가능하게 하도록 추가로 이루어지는, 시스템.

청구항 32

제1항 내지 제31항 중 어느 한 항에 있어서,

상기한 장면들에 연관된 이웃한 및/또는 인접한 환경들의 경계들로부터의 상기 사용자의 위치의 거리 또는 현재 환경 내의 상기 사용자의 위치에 연관된 다른 메트릭들 또는 미래의 환경에 대한 예측들에 기초하여 상기 서버로의 적어도 하나의 스트림의 요청을 제어하도록 이루어지는, 시스템.

청구항 33

제1항 내지 제32항 중 어느 한 항에 있어서,

각각의 오디오 엘리먼트 또는 오디오 오브젝트에 대한 정보가 서버 시스템(120)으로부터 제공되며,

상기 정보는 사운드 장면 또는 오디오 엘리먼트들이 활성화된 위치들에 관한 설명 정보를 포함하는, 시스템.

청구항 34

제1항 내지 제33항 중 어느 한 항에 있어서,

현재 또는 미래의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 및/또는 사용자의 선택에 기초하여, 하나의 장면을 재생하는 것과 적어도 2개의 장면들을 구성(compose)하거나 믹싱(mix)하거나 다중화(mux)하거나 중첩시키거나 결합시키는 것 사이에서 선택하도록 이루어지며,

상기 2개의 장면들은 상이한 이웃한 및/또는 인접한 환경들에 연관되는, 시스템.

청구항 35

제1항 내지 제34항 중 어느 한 항에 있어서,

적어도 적응 세트들을 생성 또는 사용하여,

복수의 적응 세트들이 하나의 오디오 장면과 연관되고; 그리고/또는

각각의 적응 세트를 하나의 뷰포인트, 또는 상기 하나의 오디오 장면에 관련시키는 부가적인 정보가 제공되며; 그리고/또는

- 상기 하나의 오디오 장면의 경계들에 관한 정보 및/또는,
- 하나의 적응 세트와 하나의 오디오 장면(예를 들어, 상기 오디오 장면은 3개의 적응 세트들로 캡슐화되는 3개의 스트림들로 인코딩됨) 사이의 관계에 관한 정보 및/또는,
- 상기 오디오 장면의 경계들과 상기 복수의 적응 세트들 사이의 연결에 관한 정보를 포함할 수 있는 부가적인

정보가 제공되도록 이루어지는, 시스템.

청구항 36

제1항 내지 제35항 중 어느 한 항에 있어서,

이웃한 또는 인접한 환경에 연관된 장면에 대한 스트림을 수신하고;

2개의 환경들 사이의 경계의 전환의 검출 시에 상기 이웃한 또는 인접한 환경에 대한 스트림을 디코딩 및/또는 재생하는 것을 시작하도록 이루어지는, 시스템.

청구항 37

미디어 소비 디바이스에서 재생될 비디오 및/또는 오디오 스트림들을 전달하도록 이루어진 클라이언트 및 서버로서 동작하도록 이루어진, 제1항 내지 제36항 중 어느 한 항의 시스템을 포함하는, 시스템.

청구항 38

제1항 내지 제37항 중 어느 한 항에 있어서,

상기 시스템은,

적어도 하나의 제1 오디오 장면과 연관된 적어도 하나의 오디오 스트림을 포함하는 적어도 하나의 제1 적응 세트를 요청 및/또는 수신하고;

상기 적어도 하나의 제1 오디오 장면을 포함하는 적어도 2개의 오디오 장면들과 연관된 적어도 하나의 제2 오디오 스트림을 포함하는 적어도 하나의 제2 적응 세트를 요청 및/또는 수신하며; 그리고

상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 관해 이용가능한 메타데이터, 및/또는 상기 적어도 하나의 제1 적응 세트와 상기 적어도 하나의 제1 오디오 장면의 연관 및/또는 상기 적어도 하나의 제2 적응 세트와 상기 적어도 하나의 제1 오디오 장면의 연관을 설명하는 정보에 기초하여, 상기 적어도 하나의 제1 오디오 스트림과 상기 적어도 하나의 제2 오디오 스트림을 디코딩될 새로운 오디오 스트림으로 병합하는 것을 가능하게 하도록 추가로 이루어지는, 시스템.

청구항 39

제1항 내지 제38항 중 어느 한 항에 있어서,

상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 관한 정보, 및/또는 상기 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보를 수신하고;

적어도 하나의 장면 및/또는 뷰포인트 및/또는 뷰포트 및/또는 위치 및/또는 가상 위치 및/또는 움직임 데이터 및/또는 배향에 대한 적어도 하나의 적응 세트의 연관을 설명하는 정보 및 적응 세트들의 이용가능성에 관한 정보를 수신하도록 이루어지는, 시스템.

청구항 40

제1항 내지 제39항 중 어느 한 항에 있어서,

적어도 하나의 스트림에 임베딩(embed)된 적어도 하나의 오디오 장면으로부터의 적어도 하나의 오디오 엘리먼트 및 적어도 하나의 부가적인 스트림에 임베딩된 적어도 하나의 부가적인 오디오 장면으로부터의 적어도 하나의 부가적인 오디오 엘리먼트가 재생될지를 판단하고;

긍정적인 판단의 경우, 상기 적어도 하나의 오디오 장면의 적어도 하나의 스트림에 상기 부가적인 오디오 장면의 적어도 하나의 부가적인 스트림을 병합하거나 구성(compose)하거나 다중화하거나 중첩시키거나 결합시키는 동작을 야기하도록 이루어지는, 시스템.

청구항 41

제1항 내지 제40항 중 어느 한 항에 있어서,

재생될 것으로 판단되는 오디오 장면을 조립하는 오디오 엘리먼트들을 선택하고 그리고/또는 인에이블링시키고 그리고/또는 활성화시키고 모든 선택된 오디오 스트림들의 단일 오디오 스트림으로의 병합을 가능하게 하기 위해, 상기 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여, 선택된 오디오 스트림들과 연관된 오디오 메타데이터를 조작하도록 이루어지는, 시스템.

청구항 42

가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위해 클라이언트에 오디오 및 비디오 스트림들을 전달하기 위한 서버(120)로서,

상기 비디오 및 오디오 스트림들은 미디어 소비 디바이스에서 재생되고,

상기 서버(120)는, 시각적 환경을 설명하기 위한 비디오 스트림들을 인코딩하기 위한 인코더 및/또는 상기 비디오 스트림들을 저장하기 위한 저장소를 포함하고, 상기 시각적 환경은 오디오 장면에 연관되고;

상기 서버는, 상기 클라이언트에 전달될 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 인코딩하기 위한 인코더 및/또는 상기 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 저장하기 위한 저장소를 더 포함하고, 상기 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들은 적어도 하나의 오디오 장면에 연관되고,

상기 서버는,

상기 클라이언트로부터의 요청에 기초하여, 환경에 연관된 비디오 스트림을 선택 및 전달하고;

사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에, 그리고 상기 환경에 연관된 오디오 장면에 적어도 연관된 상기 클라이언트로부터의 요청에 기초하여 오디오 스트림 및/또는 오디오 엘리먼트 및/또는 적응 세트를 선택하며;

상기 오디오 스트림을 상기 클라이언트에 전달하도록 이루어지는, 서버.

청구항 43

제42항에 있어서,

상기 스트림들은 적응 세트들로 캡슐화되며, 각각의 적응 세트는 동일한 오디오 콘텐츠의 상이한 비트레이트 및/또는 품질의 상이한 표현들에 연관된 복수의 스트림들을 포함하고,

선택된 적응 세트는 상기 클라이언트로부터의 요청에 기초하여 선택되는, 서버.

청구항 44

클라이언트 및 서버로서 동작하는 제1항 내지 제41항 중 어느 한 항의 시스템을 포함하는, 시스템.

청구항 45

제44항에 있어서,

제42항 또는 제43항의 서버를 포함하는, 시스템.

청구항 46

미디어 소비 디바이스에서 재생될 비디오 및/또는 오디오 스트림들을 수신하도록 이루어진, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 방법으로서,

사용자로서의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하는 단계, 및

오디오 스트림들로부터의 오디오 신호들을 디코딩하는 단계,

상기 사용자의 현재 뷰포트 및/또는 위치 데이터 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터 및/또는 메타데이터에 기초하여 적어도 하나의 오디오 스트림을, 서버에 요청하고 그 리고/또는 상기 서버로부터 획득하는 단계를 포함하는, 방법.

청구항 47

프로세서에 의해 실행될 경우, 상기 프로세서로 하여금 제46항의 방법을 수행하게 하는 명령들을 포함하는, 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 가상 현실(Virtual Reality; VR) 환경에서 또는 유사하게 증강 현실(Augmented Reality; AR) 또는 혼합 현실(Mixed Reality; MR) 또는 360도 비디오 환경들에서, 사용자는 일반적으로, 예를 들어 머리 장착형 디스플레이(Head Mounted Display; HMD)를 사용하여 완전한 360도 콘텐츠를 시각화하고, 헤드폰들을 통해(또한 유사하게, 자신의 위치에 의존하는 정확한 렌더링을 포함하는 라우드스피커들을 통해) 그 콘텐츠를 청취(listen)한다.

배경 기술

[0002] 간단한 사용 경우에서, 콘텐츠는, 하나의 오디오/비디오 장면(즉, 예를 들어, 360도 비디오)만이 특정한 뷰포인트에 재생되는 그러한 방식으로 작성된다. 오디오/비디오 장면은 고정된 위치(예를 들어, 사용자가 중심에 위치된 구)를 가지며, 사용자는 장면에서 이동할 수 없지만, 사용자는 단지 그의 머리를 다양한 방향들(요(yaw), 피치(pitch), 롤(roll))로 회전시킬 수 있다. 이러한 경우, 상이한 비디오 및 오디오가 사용자의 머리의 배향에 기초하여 사용자에게 재생된다(상이한 뷰포트(viewport)들이 디스플레이됨).

[0003] 비디오의 경우, 비디오 콘텐츠가 렌더링 프로세스를 설명하기 위한 메타데이터(예를 들어, 스티칭(stitching) 정보, 투영 맵핑 등)와 함께 전체 360도 장면에 대해 전달되고 현재 사용자의 뷰포트에 기초하여 선택되지만, 오디오의 경우, 콘텐츠는 전체 장면에 대해 동일하다. 메타데이터에 기초하여, 오디오 콘텐츠는 현재 사용자의 뷰포트에 적응된다(예를 들어, 오디오 오브젝트는 뷰포트/사용자 배향 정보에 기초하여 상이하게 렌더링됨). 360도 콘텐츠는 사용자가 (예를 들어, 자신의 머리 배향에 의해 또는 원격-제어 디바이스를 사용함으로써) 선택할 수 있는, 시간상으로 동일한 시점에 하나 초과 시야각으로 포함하는 임의의 타입의 콘텐츠를 지칭함을 유의해야 한다.

[0004] 더 복잡한 시나리오에서, 사용자가 VR 장면에서 이동하거나 하나의 장면으로부터 다음의 장면으로 "점프"할 수 있는 경우, 오디오 콘텐츠가 또한 변화될 수 있다(예컨대, 하나의 장면에서 가청적이지 않은 오디오 소스들은 다음 장면 - "도어(door)가 열림"에서 가청적이게 될 수 있음). 기존의 시스템들을 이용하면, 완전한 오디오 장면들이 하나의 스트림으로 그리고 필요하다면 (메인 스트림에 의존하여) 부가적인 스트림들로 인코딩될 수 있다. 그러한 시스템들은 차세대 오디오 시스템들(예를 들어, MPEG-H 3D 오디오)로 알려져 있다. 그러한 사용 경우들의 예들은 다음을 포함할 수 있다:

- [0005] ● 실시예 1: 사용자는 새로운 룸(room)으로 입장하기로 선택하고, 전체 오디오/비디오 장면이 변환된다.
- [0006] ● 실시예 2: 사용자가 VR 장면에서 이동하고, 도어를 열고 걸어 들어가서, 하나의 장면으로부터 다음 장면으로 오디오의 요구되는 전환을 암시한다.
- [0007] 이러한 시나리오를 설명하려는 목적을 위해, 공간 내의 별개의 뷰포인트들(Discret Viewpoints)의 개념은, 상이한 오디오/비디오 콘텐츠가 이용가능한 공간 내의(또는 VR 환경 내의) 별개의 위치로서 도입된다.
- [0008] "스트레이트-포워드(straight-forward)" 솔루션은, 사용자 위치/배향에 관한 재생 디바이스로부터의 피드백에 기초하여 인코딩(오디오 엘리먼트들의 수, 공간 정보 등)을 변화시키는 실시간 인코더를 갖는 것이다. 이러한 솔루션은, 예를 들어 스트리밍 환경에서, 클라이언트와 서버 사이의 매우 복잡한 통신을 암시할 것이다:
- [0009] ● (간단한 로직만을 사용하고 있는 것으로 일반적으로 가정되는) 클라이언트는 상이한 스트림들에 대한 요청들 뿐만 아니라, 사용자의 위치에 기초하여 올바른 콘텐츠의 프로세싱을 가능하게 할 인코딩 세부사항들에 관한 복잡한 정보를 전달하기 위한 발전된 메커니즘들을 요구할 것이다.
- [0010] ● 미디어 서버는 일반적으로, ("세그먼트-별(segment-wise)" 전달을 허용하는 특정 방식으로 포맷팅된) 상이한 스트림들로 미리-채워져 있으며, 서버의 메인 기능은, 이용가능한 스트림들에 관한 정보를 제공하고, 요청될 경우 스트림들의 전달을 야기하는 것이다. 재생 디바이스로부터의 피드백에 기초하여 인코딩을 허용하는 시나리오들을 가능하게 하기 위해, 미디어 서버는 복수의 라이브 미디어 인코더들과의 발전된 통신 링크들, 및 실시간으로 변화될 수 있는 모든 시그널링 정보(예를 들어, 미디어 프리젠테이션 설명)를 즉시(on the fly) 생성하기 위한 용량을 요구할 것이다.
- [0011] 그러한 시스템이 상상될 수 있지만, 그의 복잡도 및 계산 요건들은, 오늘날 이용가능하거나 또는 심지어 향후 수십년 내에 개발될 장비 및 시스템들의 기능 및 특징들을 넘어서는다.
- [0012] 대안적으로, 완전한 VR 환경("완전한 세상")을 표현하는 콘텐츠가 항상 전달될 수 있다. 이것은 문제를 해결할 것이지만, 이용가능한 통신 링크들의 용량을 넘어서는 막대한 비트레이트를 요구할 것이다.
- [0013] 이것은 실시간 환경의 경우 복잡하며, 이용가능한 시스템들을 사용하여 그러한 사용 경우들을 가능하게 하기 위해, 이러한 기능을 낮은 복잡도로 가능하게 하는 대안적인 솔루션들이 제안된다.
- [0014] 2. 용어 및 정의들
- [0015] 다음의 용어가 기술 분야에서 사용된다:
- [0016] ● **오디오 엘리먼트들**: 예를 들어, 오디오 오브젝트들, 오디오 채널들, 장면 기반 오디오(고차 앰비소닉스(Higher Order Ambisonics; HOA)) 또는 이들 모두의 임의의 조합으로서 표현될 수 있는 오디오 신호들.
- [0017] ● **관심-구역(ROI)**: 하나의 시간 순간에 사용자가 관심있는 비디오 콘텐츠의(또는 디스플레이되거나 시물레이션된 환경의) 하나의 구역. 예를 들어, 이것은 일반적으로, 구 상의 구역 또는 2D 맵으로부터의 다각형 선택일 수 있다. ROI는 특정한 목적을 위해 특정 구역을 식별하여, 고려 중인 오브젝트의 경계들을 정의한다.
- [0018] ● **사용자 위치 정보**: 위치 정보(예를 들어, x, y, z 좌표들), 배향 정보(요, 피치, 롤), 움직임의 방향 및 속도 등.
- [0019] ● **뷰포트(Viewport)**: 사용자에게 의해 현재 디스플레이 및 뷰잉(view)되는 구형 비디오의 일부.
- [0020] ● **뷰포인트(Viewpoint)**: 뷰포트의 중심 포인트.
- [0021] ● **360도 비디오**(몰입형 비디오 또는 구형 비디오로 또한 알려져 있음): 동일한 시간 순간에 하나의 방향에서 하나보다 많은 뷰(즉, 뷰포트)를 포함하는 비디오 콘텐츠를 본 명세서의 맥락에서 표현한다. 그러한 콘텐츠는, 예를 들어 무지향성(omnidirectional) 카메라 또는 카메라들의 집합을 사용하여 생성될 수 있다. 재생 동안, 뷰어(Viewer)는 뷰잉 방향의 제어가 가능하다.
- [0022] ● **미디어 프리젠테이션 설명(Media Presentation Description; MPD)**은 신택스(syntax), 예를 들어 미디어 세그먼트들에 관한 정보, 그들의 관계들 및 그들 사이에서 선택하는 데 필요한 정보를 포함하는 XML이다.
- [0023] ● **적응 세트들**은 미디어 스트림 또는 미디어 스트림들의 세트를 포함한다. 가장 간단한 경우, 하나의 적응 세트는 콘텐츠에 대한 모든 오디오 및 비디오를 포함하지만, 대역폭을 감소시키기 위해, 각각의 스트림은 상이한 적응 세트로 분할될 수 있다. 일반적인 경우는 하나의 비디오 적응 세트, 및 복수의 오디오 적응 세트들(각각의

지원되는 언어에 대해 하나씩)을 갖는 것이다. 적응 세트들은 또한 자막들 또는 임의의 메타데이터를 포함할 수 있다.

[0024] ● **표현들(Representations)**은 적응 세트가 상이한 방식으로 인코딩된 동일한 콘텐츠를 포함하게 허용한다. 대부분의 경우들에서, 표현들은 복수의 비트레이트들로 제공될 것이다. 이것은, 클라이언트들이 버퍼링을 기다리지 않으면서 재생할 수 있는 최고의 품질 콘텐츠를 클라이언트들이 요청하게 허용한다. 표현들은 또한 상이한 코덱들로 인코딩되어, 상이한 지원된 코덱들로 클라이언트들에 대해 지원할 수 있다.

[0025] 본 출원의 맥락에서, 적응 세트들의 개념들은 더 일반적으로 사용되며, 때때로 표현들을 실제로 참조한다. 또한, 미디어 스트림들(오디오/비디오 스트림들)은 일반적으로, 클라이언트(예를 들어, DASH 클라이언트)에 의해 재생되는 실제 미디어 파일들인 미디어 세그먼트들로 먼저 캡슐화된다. MPEG-4 컨테이너 포맷과 유사한 ISO 기본 미디어 파일 포맷(ISO Base Media File Format; ISO/BMFF) 또는 MPEG-2 전송 스트림(TS)과 같은 미디어 세그먼트들에 대해 다양한 포맷들이 사용될 수 있다. 미디어 세그먼트들 및 상이한 표현들/적응 세트들로의 캡슐화는 본 명세서에 설명된 방법들과 독립적이며, 방법들은 모든 다양한 옵션들에 적용된다.

[0026] 부가적으로, 본 명세서의 방법들의 설명은 DASH 서버-클라이언트 통신에 중심이 맞춰져 있지만, 방법들은 MMT, MPEG-2 TS, DASH-ROUTE, 파일 재생을 위한 파일 포맷 등과 같은 다른 전달 환경들에 대해 작동하기에 일반적으로 충분하다.

[0027] 일반적인 측면들에서, 적응 세트는 스트림에 관해 상위 계층에 있으며, (예를 들어, 위치에 연관된) 메타데이터를 포함할 수 있다. 스트림은 복수의 오디오 엘리먼트들을 포함할 수 있다. 오디오 장면은 복수의 적응 세트들의 일부로서 전달되는 복수의 스트림들에 연관될 수 있다.

선행기술문헌

비특허문헌

- [0028] (비특허문헌 0001) 3. 현재 솔루션들
- (비특허문헌 0002) 현재 솔루션들은 다음과 같다:
- (비특허문헌 0003) ISO/IEC 23008-3:2015, Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 3: 3D audio
- (비특허문헌 0004) N16950, Study of ISO/IEC DIS 23000-20 Omnidirectional Media Format
- (비특허문헌 0005) 현재 솔루션들은, 사용자가 자신의 배향을 변화시키지만 VR 환경에서 이동하지 않도록 허용하는 하나의 고정된 위치에서 독립적인 VR 경험을 제공하는 것으로 제한된다.

발명의 내용

해결하려는 과제

[0029] 대안적으로, 완전한 VR 환경("완전한 세상")을 표현하는 콘텐츠가 항상 전달될 수 있다. 이것은 문제를 해결할 것이지만, 이용가능한 통신 링크들의 용량을 넘어서는 막대한 비트레이트를 요구할 것이다.

[0030] 이것은 실시간 환경의 경우 복잡하며, 이용가능한 시스템들을 사용하여 그러한 사용 경우들을 가능하게 하기 위해, 이러한 기능을 낮은 복잡도로 가능하게 하는 대안적인 솔루션들이 제안된다.

과제의 해결 수단

[0031] 일 실시예에 따르면, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 시스템은 미디어 소비 디바이스에서 재생될 비디오 및 오디오 스트림들을 수신하도록 구성될 수 있으며, 여기서 시스템은, 사용자로서의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하도록 이루어진 적어도 하나의 미디어 비디오 디코더, 및 적어도 하나의 오디오 스트림으로부터의 오디오 신호들을 디코딩하도록 이루어진 적어도 하나의 오디오 디코더를 포함할 수 있고, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치

데이터에 적어도 기초하여, 적어도 하나의 오디오 스트림 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 서버에 요청하도록 구성될 수 있다.

[0032] 일 양상에 따르면, 시스템은, 서버로부터 적어도 하나의 오디오 스트림 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 획득하기 위해, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터를 서버에 제공하도록 구성될 수 있다.

[0033] 일 실시예는, 적어도 하나의 장면이 적어도 하나의 오디오 엘리먼트에 연관되도록 구성될 수 있으며, 각각의 오디오 엘리먼트는, 오디오 엘리먼트가 가청적인 시각적 환경 내의 위치 및/또는 영역에 연관되어, 상이한 오디오 스트림들이 장면 내의 상이한 사용자의 위치들 및/또는 뷰포트들 및/또는 머리 배향들 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 대해 제공된다.

[0034] 다른 양상에 따르면, 시스템은, 오디오 스트림의 적어도 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트가 장면 내의 현재 사용자의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 대해 재생될지 여부를 판단하도록 구성될 수 있으며, 상기 시스템은 현재 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 구성될 수 있다.

[0035] 일 양상에 따르면, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여, 오디오 스트림의 적어도 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트가 관련있게 될지 그리고/또는 가청적이게 될지를 예측적으로 판단하도록 구성될 수 있으며, 시스템은, 장면 내의 예측된 사용자의 움직임 및/또는 상호작용 이전에 특정한 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림 및/또는 적응 세트를 요청 및/또는 수신하도록 구성될 수 있고, 시스템은, 수신될 경우, 장면 내의 사용자의 움직임 및/또는 상호작용 이후 특정한 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림을 재생하도록 구성될 수 있다.

[0036] 일 실시예에서, 시스템은 장면 내의 사용자의 움직임 및/또는 상호작용 이전에 사용자의 가상 위치에서 더 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 구성될 수 있으며, 여기서 시스템은, 장면 내의 사용자의 움직임 및/또는 상호작용 이후 사용자의 가상 위치에서 더 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 구성될 수 있다.

[0037] 일 양상에 따르면, 시스템은, 적어도 하나의 오디오 엘리먼트가 적어도 하나의 장면에 연관될 수 있고, 각각의 오디오 엘리먼트는 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관될 수 있고, 시스템은, 사용자로부터 더 멀리있는 오디오 엘리먼트들보다 사용자에게 더 가까운 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질로 스트림들을 요청 및/또는 수신하도록 구성될 수 있다.

[0038] 일 양상에 따르면, 시스템은, 적어도 하나의 오디오 엘리먼트는 적어도 하나의 장면에 연관될 수 있고 적어도 하나의 오디오 엘리먼트는 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관될 수 있으며, 시스템은, 장면 내의 각각의 사용자의 가상 위치에서의 오디오 엘리먼트들의 관련성 및/또는 가청성 레벨에 기초하여 오디오 엘리먼트들에 대해 상이한 비트레이트들 및/또는 품질 레벨들로 상이한 스트림들을 요청할 수 있고, 시스템은, 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가청적인 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 오디오 스트림을 요청하고 그리고/또는 현재 사용자의 가상 위치에서 덜 관련있고 덜 가청적인 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 오디오 스트림을 요청하도록 구성될 수 있다.

[0039] 시스템의 일 실시예에서, 적어도 하나의 오디오 엘리먼트는 장면에 연관될 수 있고, 각각의 오디오 엘리먼트는 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관될 수 있으며, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터를 서버에 주기적으로 전송하여, 제1 위치에 대해, 더 높은 비트레이트 및/또는 품질의 스트림이 서버로부터 제공되고, 제2 위치에 대해, 더 낮은 비트레이트 및/또는 품질의 스트림이 서버로부터 제공되도록 구성될 수 있고, 여기서 제1 위치는 제2 위치보다 적어도 하나의 오디오 엘리먼트에 더 가깝다.

[0040] 시스템의 일 실시예에서, 복수의 장면들이 인접한 및/또는 이웃한 환경들과 같은 복수의 시각적 환경들에 대해 정의될 수 있어서, 제1의 현재 장면에 연관된 제1 스트림들이 제공되고, 제2의 추가적인 장면으로의 사용자의 전환의 경우에는, 제1 장면에 연관된 스트림들 및 제2 장면에 연관된 제2 스트림들 둘 모두가 제공된다.

[0041] 시스템의 일 실시예에서, 복수의 장면들이 제1 및 제2 시각적 환경들에 대해 정의될 수 있으며, 제1 및 제2 환경들은 인접한 및/또는 이웃한 환경들이고, 제1 장면에 연관된 제1 스트림들은, 사용자의 위치 또는 가상 위치

가 제1 장면에 연관된 제1 환경에 있는 경우 제1 장면의 재생을 위해 서버로부터 제공되고, 제2 장면에 연관된 제2 스트림들은, 사용자의 위치 또는 가상 위치가 제2 장면에 연관된 제2 환경에 있는 경우 제2 장면의 재생을 위해 서버로부터 제공되며, 제1 장면에 연관된 제1 스트림들 및 제2 장면에 연관된 제2 스트림들 둘 모두는 사용자의 위치 또는 가상 위치가 제1 장면과 제2 장면 사이의 전환 위치에 있는 경우 제공된다.

[0042] 시스템의 일 실시예에서, 복수의 장면들은 인접한 및/또는 이웃한 환경들인 제1 및 제2 시각적 환경들에 대해 정의되도록 구성될 수 있고, 시스템은, 사용자의 가상 위치가 제1 환경에 있는 경우 제1 장면의 재생을 위해 제1 환경에 연관된 제1 장면에 연관된 제1 스트림들을 요청 및/또는 수신하도록 이루어지고, 시스템은, 사용자의 가상 위치가 제2 환경에 있는 경우 제2 장면의 재생을 위해 제2 환경에 연관된 제2 장면에 연관된 제2 스트림들을 요청 및/또는 수신하도록 구성될 수 있으며, 시스템은, 사용자의 가상 위치가 제1 환경과 제2 환경 사이의 전환 위치에 있는 경우 제1 장면에 연관된 제1 스트림들 및 제2 장면에 연관된 제2 스트림들 둘 모두를 요청 및/또는 수신하도록 구성될 수 있다.

[0043] 일 양상에 따르면, 시스템은, 사용자가 제1 장면에 연관된 제1 환경에 있는 경우 제1 장면에 연관된 제1 스트림들이 더 높은 비트레이트 및/또는 품질로 획득되는 반면, 사용자가 제1 장면으로부터 제2 장면으로의 전환 위치의 시작부에 있는 경우 제2 환경에 연관된 제2 장면에 연관된 제2 스트림들이 더 낮은 비트레이트 및/또는 품질로 획득되며, 사용자가 제1 장면으로부터 제2 장면으로의 전환 위치의 끝에 있는 경우 제1 장면에 연관된 제1 스트림들이 더 낮은 비트레이트 및/또는 품질로 획득되고, 제2 장면에 연관된 제2 스트림들이 더 높은 비트레이트 및/또는 품질로 획득되도록 구성될 수 있고, 더 낮은 비트레이트 및/또는 품질은 더 높은 비트레이트 및/또는 품질보다 낮다.

[0044] 일 양상에 따르면, 시스템은 복수의 장면들이 인접한 및/또는 이웃한 환경들과 같은 복수의 환경들에 대해 정의될 수 있도록 구성될 수 있고, 시스템은, 제1의 현재 환경에 연관된 제1 현재 장면에 연관된 스트림들을 획득할 수 있고, 장면의 경계로부터의 사용자의 위치 또는 가상 위치의 거리가 미리 결정된 임계치 미만인 경우, 시스템은, 제2 장면에 연관된 제2의 인접한 및/또는 이웃한 환경에 연관된 오디오 스트림들을 추가로 획득할 수 있다.

[0045] 일 양상에 따르면, 시스템은 복수의 장면들이 복수의 시각적 환경들에 대해 정의될 수 있도록 구성될 수 있으며, 시스템은, 더 높은 비트레이트 및/또는 품질로 현재 장면에 연관된 스트림들을 그리고 더 낮은 비트레이트 및/또는 품질로 제2 장면에 연관된 스트림들을 요청 및/또는 획득하며, 더 낮은 비트레이트 및/또는 품질은 더 높은 비트레이트 및/또는 품질보다 낮다.

[0046] 일 양상에 따르면, 시스템은 복수의 N개의 오디오 엘리먼트들이 정의될 수 있도록 구성될 수 있으며, N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 작은 경우 N개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림을 시스템에 제공하거나 N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 큰 경우 M개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림을 시스템에 제공하기 위해 이들 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 큰 경우, N개의 오디오 엘리먼트들은 N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들($M < N$)을 획득하도록 프로세싱되도록 구성된다.

[0047] 일 양상에 따르면, 시스템은 적어도 하나의 시각적 환경 장면이 적어도 하나의 복수의 N개의 오디오 엘리먼트들($N \geq 2$)에 연관되도록 구성될 수 있으며, 각각의 오디오 엘리먼트는 시각적 환경 내의 위치 및/또는 영역에 연관되고, 복수의 N개의 오디오 엘리먼트들 중 적어도 하나는 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고, 복수의 N개의 오디오 엘리먼트들 중 적어도 하나는 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고, 적어도 하나의 표현은, N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들($M < N$)을 획득하도록 N개의 오디오 엘리먼트들을 프로세싱함으로써 획득되고, 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가청적인 경우 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 표현을 요청하도록 구성될 수 있고, 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 덜 관련있고 그리고/또는 덜 가청적인 경우 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 표현을 요청하도록 구성될 수 있다.

[0048] 일 양상에 따르면, 시스템은, 사용자의 거리 및/또는 관련성 및/또는 가청성 레벨 및/또는 각도 배향이 미리 결정된 임계치보다 낮은 경우, 상이한 스트림들이 상이한 오디오 엘리먼트들에 대해 획득되도록 구성될 수 있다.

- [0049] 일 실시예에서, 시스템은, 장면 내의 사용자의 배향 및/또는 사용자의 움직임 방향 및/또는 사용자의 상호작용들에 기초하여 스트림들을 요청 및/또는 획득하도록 구성될 수 있다.
- [0050] 시스템의 일 실시예에서, 뷰포트는 위치 및/또는 가상 위치 및/또는 움직임 데이터 및/또는 머리에 연관될 수 있다.
- [0051] 일 양상에 따르면, 시스템은 상이한 오디오 엘리먼트들이 상이한 뷰포트들에서 제공되도록 구성될 수 있으며, 시스템은, 하나의 제1 오디오 엘리먼트가 뷰포트 내에 있는 경우, 뷰포트 내에 있지 않은 제2 오디오 엘리먼트보다 더 높은 비트레이트로 제1 오디오 엘리먼트를 요청 및/또는 수신하도록 구성될 수 있다.
- [0052] 일 양상에 따르면, 시스템은 제1 오디오 스트림들 및 제2 오디오 스트림들을 요청 및/또는 수신하도록 구성될 수 있으며, 제1 오디오 스트림들 내의 제1 오디오 엘리먼트들은 제2 오디오 스트림들 내의 제2 오디오 엘리먼트들보다 더 관련있고 그리고/또는 더 가치적이고, 제1 오디오 스트림들은 제2 오디오 스트림들의 비트레이트 및/또는 품질보다 더 높은 비트레이트 및/또는 품질로 요청 및/또는 수신된다.
- [0053] 일 양상에 따르면, 시스템은 적어도 2개의 시각적 환경 장면들이 정의되도록 구성될 수 있으며, 적어도 하나의 제1 및 제2 오디오 엘리먼트들은 제1 시각적 환경에 연관된 제1 장면에 연관되고, 적어도 하나의 제3 오디오 엘리먼트는 제2 시각적 환경에 연관된 제2 장면에 연관되고, 시스템은 적어도 하나의 제2 오디오 엘리먼트가 제2 시각적 환경 장면과 부가적으로 연관된다는 것을 설명하는 메타데이터를 획득하도록 구성될 수 있고, 시스템은, 사용자의 가상 위치가 제1 시각적 환경에 있는 경우 적어도 제1 및 제2 오디오 엘리먼트들을 요청 및/또는 수신하도록 구성될 수 있고, 시스템은, 사용자의 가상 위치가 제2 시각적 환경 장면에 있는 경우 적어도 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 구성될 수 있으며, 시스템은, 사용자의 가상 위치가 제1 시각적 환경 장면과 제2 시각적 환경 장면 사이의 전환에 있는 경우 적어도 제1 및 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 구성될 수 있다.
- [0054] 시스템의 일 실시예는, 적어도 하나의 제1 오디오 엘리먼트가 적어도 하나의 오디오 스트림 및/또는 적응 세트에 제공되고, 적어도 하나의 제2 오디오 엘리먼트가 적어도 하나의 제2 오디오 스트림 및/또는 적응 세트에 제공되며, 적어도 하나의 제3 오디오 엘리먼트가 적어도 하나의 제3 오디오 스트림 및/또는 적응 세트에 제공되도록 구성될 수 있고, 적어도 제1 시각적 환경 장면은 적어도 제1 및 제2 오디오 스트림들 및/또는 적응 세트들을 요구하는 완전한 장면으로서 메타데이터에 의해 설명되고, 제2 시각적 환경 장면은 적어도 제1 시각적 환경 장면과 연관된 적어도 제3 오디오 스트림들 및/또는 적응 세트들 및 적어도 제2 오디오 스트림 및/또는 적응 세트들을 요구하는 불완전한 장면으로서 메타데이터에 의해 설명되며, 시스템은, 사용자의 가상 위치가 제2 시각적 환경에 있는 경우, 제1 시각적 환경에 속하는 제2 오디오 스트림 및 제2 시각적 환경과 연관된 제3 오디오 스트림을 새로운 단일 스트림으로 병합하는 것을 허용하기 위해 메타데이터를 조작하도록 이루어진 메타데이터 프로세서를 포함한다.
- [0055] 일 양상에 따르면, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 메타데이터를 조작하도록 구성된 메타데이터 프로세서를 포함한다.
- [0056] 일 양상에 따르면, 메타데이터 프로세서는, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 인에이블링 및/또는 디스에이블링시키도록 구성될 수 있으며, 메타데이터 프로세서는, 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 오디오 엘리먼트가 더 이상 재생되지 않을 것이라고 시스템이 판단하는 경우, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 디스에이블링시키도록 구성될 수 있고, 메타데이터 프로세서는, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 오디오 엘리먼트가 재생될 것이라고 시스템이 판단하는 경우, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 인에이블링시키도록 구성될 수 있다.
- [0057] 일 양상에 따르면, 시스템은 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치에 기초하여 선택된 오디오 엘리먼트들의 디코딩을 디스에이블링시키도록 구성될 수 있다.
- [0058] 일 양상에 따르면, 시스템은, 현재 오디오 장면에 연관된 적어도 하나의 제1 오디오 스트림을 이웃한, 인접한 및/또는 미래의 오디오 장면에 연관된 적어도 하나의 스트림에 병합하도록 구성될 수 있다.

- [0059] 일 양상에 따르면, 시스템은 통계 또는 집계된 데이터에 연관된 요청을 서버에 송신하기 위해 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 대한 통계 또는 집계된 데이터를 획득 및/또는 수집하도록 구성될 수 있다.
- [0060] 일 양상에 따르면, 시스템은 적어도 하나의 스트림에 연관된 메타데이터에 기초하여 그리고 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 기초하여 적어도 하나의 스트림의 디코딩 및/또는 재생을 비활성화시키도록 구성될 수 있다.
- [0061] 일 양상에 따르면, 시스템은, 재생될 오디오 장면을 구성(compose)하는 오디오 엘리먼트들을 선택하고 그리고/또는 인에이블링시키고 그리고/또는 활성화시키고; 그리고/또는 모든 선택된 오디오 스트림들의 단일 오디오 스트림으로의 병합을 가능하게 하기 위해, 사용자의 현재 또는 추정된 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여, 선택된 오디오 스트림들의 그룹과 연관된 메타데이터를 조작하도록 구성될 수 있다.
- [0062] 일 양상에 따르면, 시스템은, 상이한 장면들에 연관된 이웃한 및/또는 인접한 환경들의 경계들로부터의 사용자의 위치의 거리 또는 현재 환경 내의 사용자의 위치에 연관된 다른 메트릭들 또는 미래의 환경에 대한 예측들에 기초하여 서버로의 적어도 하나의 스트림의 요청을 제어하도록 구성될 수 있다.
- [0063] 시스템의 일 양상에 따르면, 각각의 오디오 엘리먼트 또는 오디오 오브젝트에 대한 정보가 서버 시스템으로부터 제공될 수 있으며, 정보는 사운드 장면 또는 오디오 엘리먼트들이 활성화된 위치들에 관한 설명 정보를 포함한다.
- [0064] 일 양상에 따르면, 시스템은, 현재 또는 미래의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 및/또는 사용자의 선택에 기초하여, 하나의 장면을 재생하는 것과 적어도 2개의 장면들을 구성(compose)하거나 혼합(mix)하거나 다중화(mux)하거나 중첩(superpose)시키거나 결합(combine)시키는 것 사이에서 선택하도록 구성될 수 있으며, 2개의 장면들은 상이한 이웃한 및/또는 인접한 환경들에 연관된다.
- [0065] 일 양상에 따르면, 시스템은, 적어도 적응 세트들을 생성 또는 사용하여, 복수의 적응 세트들은 하나의 오디오 장면과 연관되고; 그리고/또는 각각의 적응 세트를 하나의 뷰포인트(viewpoint), 또는 하나의 오디오 장면에 관련시키는 부가적인 정보가 제공되며; 그리고/또는 하나의 오디오 장면의 경계들에 관한 정보 및/또는 하나의 적응 세트와 하나의 오디오 장면(예를 들어, 오디오 장면은 3개의 적응 세트들로 캡슐화되는 3개의 스트림들로 인코딩됨) 사이의 관계에 관한 정보 및/또는 오디오 장면의 경계들과 복수의 적응 세트들 사이의 연결에 관한 정보를 포함할 수 있는 부가적인 정보가 제공되도록 구성될 수 있다.
- [0066] 일 양상에 따르면, 시스템은, 이웃한 또는 인접한 환경에 연관된 장면에 대한 스트림을 수신하고; 2개의 환경들 사이의 경계의 전환의 검출 시에 이웃한 또는 인접한 환경에 대한 스트림을 디코딩 및/또는 재생하는 것을 시작하도록 구성될 수 있다.
- [0067] 일 양상에 따르면, 시스템은, 미디어 소비 디바이스에서 재생될 비디오 및/또는 오디오 스트림들을 전달하도록 구성된 클라이언트 및 서버로서 동작하도록 구성될 수 있다.
- [0068] 일 양상에 따르면, 시스템은, 적어도 하나의 제1 오디오 장면과 연관된 적어도 하나의 오디오 스트림을 포함하는 적어도 하나의 제1 적응 세트를 요청 및/또는 수신하고; 적어도 하나의 제1 오디오 장면을 포함하는 적어도 2개의 오디오 장면들과 연관된 적어도 하나의 제2 오디오 스트림을 포함하는 적어도 하나의 제2 적응 세트를 요청 및/또는 수신하며; 그리고 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 관해 이용가능한 메타데이터 및/또는 적어도 하나의 제1 적응 세트와 적어도 하나의 제1 오디오 장면의 연관 및/또는 적어도 하나의 제2 적응 세트와 적어도 하나의 제1 오디오 장면의 연관을 설명하는 정보에 기초하여, 적어도 하나의 제1 오디오 스트림과 적어도 하나의 제2 오디오 스트림을 디코딩될 새로운 오디오 스트림으로 병합하는 것을 가능하게 하도록 구성될 수 있다.
- [0069] 일 양상에 따르면, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 관한 정보 및/또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보를 수신하고; 그리고 적어도 하나의 적응 세트와 적어도 하나의 장면 및/또는 뷰포인트 및/또는 뷰포트 및/또는 위치 및/또는 가상 위치 및/또는 움직임 데이터 및/또는 배향의 연관을 설명하는 정보 및 적응 세트들의 이용가능성에 관한 정보를 수신하도록 구성될 수 있다.
- [0070] 일 양상에 따르면, 시스템은, 적어도 하나의 스트림에 임베딩(embed)된 적어도 하나의 오디오 장면으로부터의 적어도 하나의 오디오 엘리먼트 및 적어도 하나의 부가적인 스트림에 임베딩된 적어도 하나의 부가적인 오디오

장면으로부터의 적어도 하나의 부가적인 오디오 엘리먼트가 재생될 것인지를 판단하고; 긍정적인 판단의 경우, 적어도 하나의 오디오 장면의 적어도 하나의 스트림과 부가적인 오디오 장면의 적어도 하나의 부가적인 스트림을 병합하거나 합성하거나 다중화하거나 중첩시키거나 결합시키는 동작을 야기하도록 구성될 수 있다.

[0071] 일 양상에 따르면, 시스템은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여, 선택된 오디오 스트림들과 연관된 오디오 메타데이터를 조작하여, 재생될 것으로 판단되는 오디오 장면을 조립하는 오디오 엘리먼트들을 선택하고 그리고/또는 인에이블링시키고 그리고/또는 활성화시키고; 그리고 모든 선택된 오디오 스트림들의 단일 오디오 스트림으로의 병합을 가능하게 하도록 구성될 수 있다.

[0072] 일 양상에 따르면, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위해 클라이언트에 오디오 및 비디오 스트림들을 전달하기 위한 서버가 제공될 수 있으며, 비디오 및 오디오 스트림들은 미디어 소비 디바이스에서 재생되고, 서버는, 시각적 환경을 설명하기 위한 비디오 스트림들을 인코딩하기 위한 인코더 및/또는 그 비디오 스트림들을 저장하기 위한 저장소를 포함할 수 있고, 시각적 환경은 오디오 장면과 연관되고; 서버는, 클라이언트에 전달될 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 인코딩하기 위한 인코더 및/또는 그 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 저장하기 위한 저장소를 더 포함할 수 있고, 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들은 적어도 하나의 오디오 장면과 연관되고, 서버는, 클라이언트로부터의 요청에 기초하여 환경에 연관되는 비디오 스트림을 선택 및 전달하고; 클라이언트로부터의 요청 - 그 요청은 적어도 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에, 그리고 환경에 연관된 오디오 장면과 연관됨 - 에 기초하여 오디오 스트림 및/또는 오디오 엘리먼트 및/또는 적응 세트를 선택하며; 그리고 오디오 스트림을 클라이언트에 전달하도록 이루어진다.

[0073] 일 양상에 따르면, 스트림들은 적응 세트들로 캡슐화될 수 있으며, 각각의 적응 세트는 동일한 오디오 콘텐츠의 상이한 비트레이트 및/또는 품질의 상이한 표현들에 연관된 복수의 스트림들을 포함하고, 선택된 적응 세트는 클라이언트로부터의 요청에 기초하여 선택된다.

[0074] 일 양상에 따르면, 시스템은 클라이언트 및 서버로서 동작할 수 있다.

[0075] 일 양상에 따르면, 시스템은 서버를 포함할 수 있다.

[0076] 일 양상에 따르면, 미디어 소비 디바이스(예를 들어, 재생 디바이스)에서 재생될 비디오 및/또는 오디오 스트림들을 수신하도록 구성된, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 방법이 제공될 수 있으며, 그 방법은, 사용자로서의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하는 단계; 및 오디오 스트림들로부터의 오디오 신호들을 디코딩하는 단계, 사용자의 현재 뷰포트 및/또는 위치 데이터 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터 및/또는 메타데이터에 기초하여 적어도 하나의 오디오 스트림을, 서버에 요청하고 그리고/또는 서버로부터 획득하는 단계를 포함한다.

[0077] 일 양상에 따르면, 프로세서에 의해 실행될 경우, 프로세서로 하여금 위의 방법을 수행하게 하는 명령들을 포함하는 컴퓨터 프로그램이 제공될 수 있다.

발명의 효과

[0078] 본 발명에 따르면, 본 발명은 별개의 수의 시각적 환경들 및 오디오 장면들의 전통적인 접근법을 넘어서, 상이한 환경들/장면들의 점진적인 표현을 허용하여, 사용자에게 더 현실적인 경험을 제공할 수 있다.

도면의 간단한 설명

[0079] 도 1a 내지 도 1h는 본 발명의 예들을 도시한다.

도 2 내지 도 6은 본 발명의 시나리오들을 도시한다.

도 7a 내지 도 8b는 본 발명의 방법들을 도시한다.

발명을 실시하기 위한 구체적인 내용

[0080] 본 명세서의 아래(예를 들어, 다음의 도 1a)에서, 본 발명의 양상들에 따른 시스템들의 개시된 예들이

존재한다.

- [0081] (아래에서 개시되는 상이한 예들에 의해 구현될 수 있는) 본 발명의 시스템의 예들은 총괄하여 102로 표시된다. 시스템(102)은, 예를 들어 그것이 사용자로의 오디오 장면들 및/또는 시각적 환경들의 표현을 위한 오디오 및/또는 비디오 스트림들을 서버 시스템(예를 들어, 120)으로부터 획득할 수 있으므로 클라이언트 시스템일 수 있다. 클라이언트 시스템(102)은 또한, 예를 들어 오디오 및/또는 비디오 스트림들에 관한 사이드(side) 및/또는 보조 정보를 제공하는 메타데이터를 서버 시스템(120)으로부터 수신할 수 있다.
- [0082] 시스템(102)은 사용자에게 오디오 및/또는 비디오 신호들을 실제로 재생하는 미디어 소비 디바이스(MCD)에 연관될 수 있다(또는 일부 예들에서는 MCD를 포함할 수 있음). 일부 예들에서, 사용자는 MCD를 착용할 수 있다.
- [0083] 시스템(102)은 서버 시스템(120)에 대한 요청들을 수행할 수 있으며, 요청들은 적어도 하나의 사용자의 현재 뷰포트 및/또는 머리 배향(예를 들어, 각도 배향) 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 연관된다(수 개의 메트릭들이 제공될 수 있음). 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)는 MCD로부터 클라이언트 시스템(102)으로 피드백으로 제공될 수 있으며, 차례로, 클라이언트 시스템(102)은 이러한 피드백에 기초하여 요청을 서버 시스템(120)에 제공할 수 있다.
- [0084] 일부 경우들에서, (112로 표시된) 요청은 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)(또는 이들의 표시 또는 프로세싱된 버전)를 포함할 수 있다. 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 기초하여, 서버 시스템(120)은 필요한 오디오 및/또는 비디오 스트림들 및/또는 메타데이터를 제공할 것이다. 이러한 경우, 서버 시스템(120)은 (예를 들어, 가상 환경 내의) 사용자의 위치의 지식을 가질 수 있고, 정확한 스트림들을 사용자의 위치들에 연관시킬 수 있다.
- [0085] 다른 경우들에서, 클라이언트 시스템(102)으로부터의 요청(112)은 특정한 오디오 및/또는 비디오 스트림들의 명시적인 요청들을 포함할 수 있다. 이러한 경우, 요청(112)은 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 기초할 수 있다. 클라이언트 시스템(102)이 필요한 스트림들을 갖지 않거나 내부에 저장하지 않았더라도, 클라이언트 시스템(102)은 사용자에게 렌더링되어야 하는 오디오 및 비디오 신호들의 지식을 갖는다. 예들에서, 클라이언트 시스템(102)은 서버 시스템(120)에서 특정한 스트림들을 다룰 수 있다.
- [0086] 클라이언트 시스템(102)은 미디어 소비 디바이스에서 재생될 비디오 및 오디오 스트림들을 수신하도록 이루어진, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 시스템일 수 있으며,
- [0087] 여기서, 시스템(102)은,
- [0088] 사용자로의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하도록 이루어진 적어도 하나의 미디어 비디오 디코더, 및
- [0089] 적어도 하나의 오디오 스트림(106)으로부터의 오디오 신호들(108)을 디코딩하도록 이루어진 적어도 하나의 오디오 디코더(104)를 포함하고,
- [0090] 시스템(102)은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 적어도 하나의 오디오 스트림(106) 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 서버(120)에 요청(112)하도록 이루어진다.
- [0091] VR, AR, MR 환경들에서, 사용자(140)가 특정한 환경(예를 들어, 특정한 룸)에 있다고 의미될 수 있음을 유의한다. 환경은, 예를 들어 서버-측(서버 시스템(120)을 반드시 포함할 필요는 없지만, 서버(120)의 저장소에 나중에 저장되는 비디오 스트림들을 이전에 인코딩했던 상이한 인코더를 포함할 수 있는 서버 시스템(120)의 측)에서 인코딩되는 비디오 신호들을 이용하여 설명된다. 각각의 순간에, 일부 예들에서, 사용자는 일부 비디오 신호들(예를 들어, 뷰포트)만을 즐길 수 있다.
- [0092] 일반적으로, 각각의 환경은 특정한 오디오 장면에 연관될 수 있다. 오디오 장면은, 특정한 환경에서 그리고 특정한 시간 기간에 걸쳐 사용자에게 재생될 모든 사운드들의 집합으로서 이해될 수 있다.
- [0093] 전통적으로, 환경들은 별개의 수인 것으로 이해되었다. 따라서, 환경들의 수는 유한한 것으로 이해되었다. 동일한 이유들로, 오디오 장면들의 수는 유한한 것으로 이해되었다. 따라서, 종래 기술에서, VR, AR, MR 시스템들은

다음과 같이 설계되었다:

- [0094] - 사용자는 각각의 시간에 하나의 단일 환경에 있도록 의도되며; 따라서, 각각의 환경에 대해,
- [0095] ○ 클라이언트 시스템(102)은 단일 환경에 연관된 비디오 스트림들만을 서버 시스템(120)에 요청하고;
- [0096] ○ 클라이언트 시스템(102)은 단일 장면에 연관된 오디오 스트림들만을 서버 시스템(120)에 요청한다.
- [0097] 이러한 접근법은 불편함들을 가져왔다.
- [0098] 예를 들어, 모든 오디오 스트림들은 각각의 장면/환경에 대해 클라이언트 시스템(102)으로 모두 함께 전달되어야 하며, 사용자가 상이한 환경들로 이동할 경우(예를 들어, 사용자가 도어를 통과하여, 그에 따라 환경들/장면들의 송신을 암시할 경우) 완전히 새로운 오디오 스트림들이 전달될 필요가 있다.
- [0099] 추가로, 부자연스러운 경험이 일부 경우들에서 야기되었으며: 예를 들어, 사용자가 벽(예를 들어, 가상 룸의 가상 벽)에 가까이 있는 경우, 사용자는 벽의 다른 측으로부터 나오는 사운드들을 경험해야 한다. 그러나, 이러한 경험은 기존의 환경들에서는 불가능하며: 현재 장면에 연관된 오디오 스트림들의 집합은 인접한 환경들/장면들에 연관된 어떠한 스트림도 확실히 포함하지 않는다.
- [0100] 반면에, 사용자의 경험은 일반적으로, 오디오 스트림의 비트레이트가 증가될 경우 개선된다. 이것은 추가적인 이슈들을 야기할 수 있으며: 비트레이트가 높을수록, 서버 시스템이 클라이언트 시스템(102)에 전달할 필요가 있는 페이로드가 많아진다. 예를 들어, 오디오 장면이 (오디오 엘리먼트들로서 전달되는) 복수의 오디오 소스들(이들 중 일부는 사용자 위치 부근에 위치되고, 다른 것들은 사용자로부터 멀리 떨어져 위치됨)을 포함할 경우, 멀리 떨어져 위치된 사운드 소스들은 덜 가청적일 것이다. 따라서, 모든 오디오 엘리먼트들을 동일한 비트레이트 또는 품질 레벨로 전달하는 것은 매우 높은 비트레이트들을 유발할 수 있다. 이것은 비-효율적인 오디오 스트림 전달을 암시한다. 서버 시스템(120)이 최고의 가능한 비트레이트로 오디오 스트림들을 전달하면, 전체 오디오 장면에 비해 낮은 가청성 레벨 또는 낮은 관련성을 갖는 사운드들이 그럼에도 불구하고, 사용자에게 더 가깝게 생성된 관련 사운드들과 유사하게 높은 비트레이트를 요구할 것이므로, 비효율적인 전달이 야기된다. 따라서, 하나의 장면의 모든 오디오 스트림들이 최고의 비트레이트로 전달되면, 서버 시스템(120)과 클라이언트 시스템(102) 사이의 통신은 페이로드를 불필요하게 증가시킬 것이다. 하나의 장면의 모든 오디오 스트림들이 더 낮은 비트레이트로 전달되면, 사용자의 경험은 만족스럽지 않을 것이다.
- [0101] 통신의 문제들은 위에서 논의된 불편함을 악화시키며: 사용자가 도어를 통과할 경우, 사용자는 환경/장면을 즉시 변화시켜야 하는데, 이는, 서버 시스템(120)이 즉시 모든 스트림들을 클라이언트 시스템(102)에 제공해야 한다는 것을 요구할 것이다.
- [0102] 따라서, 전통적으로는, 위에서 논의된 문제들을 해결하는 것이 가능하지 않았다.
- [0103] 그러나, 본 발명을 이용하면, 이들 문제들을 해결하는 것이 가능하며: 클라이언트 시스템(102)은, 사용자의 현재 뷰포트 및/또는 머리 방향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 또한 기반할 수 있는(그리고 환경/장면에만 기초하지는 않는) 요청을 서버 시스템(120)에 제공한다. 따라서, 서버 시스템(120)은 각각의 순간에 대해, 예를 들어 각각의 사용자의 위치에 대해 렌더링될 오디오 스트림들을 제공할 수 있다.
- [0104] 예를 들어, 사용자가 결코 벽에 가까이 가지 않으면, 클라이언트 시스템(102)이 이웃한 환경의 스트림들을 요청할 필요성이 존재하지 않는다(예를 들어, 그 스트림들은, 사용자가 벽에 접근할 경우에만 클라이언트 시스템(102)에 의해 요청될 수 있음). 게다가, 벽 외부로부터 나오는 스트림들은, 그들이 낮은 볼륨으로 들릴 수 있으므로, 감소된 비트레이트를 가질 수 있다. 특히, 더 관련있는 스트림들(예를 들어, 현재 환경 내의 오디오 오브젝트들로부터 나오는 스트림들)은 (덜 관련있는 스트림들이 더 낮은 비트레이트 및/또는 품질 레벨로 이루어지므로, 그에 따라 더 관련있는 스트림들에 대한 자유 대역(free band)을 남긴다는 사실의 결과로서) 최고의 비트레이트 및/또는 최고의 품질 레벨로 서버 시스템(120)에 의해 클라이언트 시스템(102)에 전달될 수 있다.
- [0105] 더 낮은 품질 레벨은, 예를 들어 비트레이트를 감소시킴으로써 또는 송신될 요구되는 데이터가 감소되는 반면, 오디오 신호 당 사용되는 비트레이트가 일정하게 유지되는 그러한 방식으로 오디오 엘리먼트들을 프로세싱함으로써 획득될 수 있다. 예를 들어, 복수의 10개의 오디오 오브젝트들이 모두 사용자로부터 멀리 떨어진 상이한 위치들에 위치되면, 이들 오브젝트들은 사용자 위치에 기초하여 더 적은 수의 신호들로 믹싱될 수 있으며:
- [0106] - 사용자 위치로부터 매우 멀리 떨어진 위치들(예를 들어, 제1 임계치보다 높음)에서, 오브젝트들은 2개의 신호

들로 믹싱되고(그들의 공간 위치 및 시멘틱(semantic)에 기초하여, 다른 수들이 가능함) 2개의 "가상 오브젝트들"로서 전달된다.

- [0107] - 사용자 위치에 더 가까운 위치들(예를 들어, 제1 임계치보다는 낮지만, 제1 임계치보다 작은 제2 임계치보다 높음)에서, 오브젝트들은 (그들의 공간 위치 및 시멘틱에 기초하여) 5개의 신호들로 믹싱되고 5개(다른 수들이 가능함)의 "가상 오브젝트들"로서 전달된다.
- [0108] - 사용자의 위치들에 매우 가까운 위치들(제1 및 제2 임계치보다 낮음)에서, 10개의 오브젝트들은 최고의 품질을 제공받는 10개의 오디오 신호들로서 전달된다.
- [0109] 최고의 품질의 경우, 오디오 신호들 모두가 매우 중요하고 가청적인 것으로 고려될 수 있지만, 사용자는 각각의 오브젝트를 개별적으로 로컬화시키는 것이 가능할 수 있다. 멀리 떨어진 위치들에서의 더 낮은 품질 레벨들의 경우, 오디오 오브젝트들 중 일부는 덜 관련있거나 덜 가청적일 수 있으며, 따라서 사용자는 아무리 해도 공간에서 오디오 신호들을 개별적으로 로컬화할 수 없을 것이고, 따라서 이들 오디오 신호들의 전달을 위해 품질 레벨을 감소시키는 것은 사용자에게 대한 경험의 품질의 어떠한 감소도 유발하지 않을 것이다.
- [0110] 다른 예는 사용자가 도어를 넘어갈 경우이며: 전환 위치에서(예를 들어, 2개의 상이한 환경들/장면들 사이의 경계에서), 서버 시스템(120)은 장면들/환경 둘 모두의 스트림들 둘 모두를 제공하지만 더 낮은 비트레이트들로 제공할 것이다. 이것은, 사용자가 2개의 상이한 환경들로부터 나오는 사운드들을 경험할 것이고(사운드들은 상이한 장면들/환경들에 원래 연관된 상이한 오디오 스트림들로부터 함께 병합될 수 있음) 각각의 사운드 소스(또는 오디오 엘리먼트)의 최고의 품질 레벨에 대한 필요성이 발생하지 않기 때문이다.
- [0111] 위의 관점에서, 본 발명은 별개의 수의 시각적 환경들 및 오디오 장면들의 전통적인 접근법을 넘어서는 것을 허용하지만, 상이한 환경들/장면들의 점진적인 표현을 허용하여, 사용자에게 더 현실적인 경험을 제공할 수 있다.
- [0112] 본 명세서의 아래에서, 각각의 시각적 환경(예를 들어, 가상 환경)이 오디오 장면에 연관된다고 고려된다(환경들의 속성들은 또한 장면의 속성들일 수 있음). 각각의 환경/장면은, 예를 들어 (가상 기하학적 좌표계일 수 있는) 기하학적 좌표계에 연관될 수 있다. 환경/장면은 경계들을 가질 수 있으므로, 사용자의 위치(예를 들어, 가상 위치)가 경계들을 넘어갈 경우, 상이한 환경/장면에 도달한다. 경계들은 사용된 좌표계에 기반할 수 있다. 환경은, 환경/장면의 일부 특정한 좌표들에 위치될 수 있는 오디오 오브젝트들(오디오 엘리먼트들, 오디오 소스들)을 포함할 수 있다. 예를 들어, 오디오 오브젝트들(오디오 엘리먼트들, 사운드 소스들)에 관한 사용자의 상대적인 위치 및/또는 배향에 관해, 클라이언트 시스템(102)은 상이한 스트림들을 요청할 수 있고 그리고/또는 서버 시스템(120)은 (예를 들어, 거리 및/또는 배향에 따라 더 높은/더 낮은 비트레이트들 및/또는 품질 레벨들로) 상이한 스트림들을 제공할 수 있다.
- [0113] 더 일반적으로, 클라이언트 시스템(102)은 상이한 스트림들(예를 들어, 상이한 비트레이트들 및/또는 품질 레벨들의 동일한 사운드들의 상이한 표현들)을 그들의 가청성 및/또는 관련성에 기초하여 서버 시스템(120)에 요청하고 그리고/또는 서버 시스템(120)으로부터 획득할 수 있다. 가청성 및/또는 관련성은, 예를 들어 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 적어도 기초하여 결정될 수 있다.
- [0114] 수 개의 예들에서, 상이한 스트림들을 병합할 가능성이 있다. 수 개의 경우들에서, 적어도 2개의 장면들을 구성(compose)하거나 믹싱하거나 다중화하거나 중첩시키거나 결합시킬 수 있다. 예를 들어, 믹서 및/또는 렌더러(예를 들어, 적어도 하나의 오디오 스트림을 각각 디코딩하는 복수의 디코더들의 다운스트림(downstream)에서 사용될 수 있음)를 사용하거나, 또는 예를 들어 스트림들의 디코딩의 업스트림(upstream)에서 스트림 다중화 동작을 수행할 수 있다. 다른 경우들에서, 상이한 스트림들을 디코딩하고 상이한 라우드스피커 셋업들을 이용하여 그들을 렌더링할 수 있다.
- [0115] 본 발명이 반드시 시각적 환경 및 오디오 장면의 개념을 거부하는 것은 아님을 유의해야 한다. 특히, 본 발명에 따르면, 특정한 장면/환경에 연관된 오디오 및 비디오 스트림들은, 사용자가 환경/장면에 진입할 경우 서버 시스템(120)으로부터 클라이언트 시스템(102)으로 전달될 수 있다. 그럼에도 불구하고, 동일한 환경/장면 내에서, 상이한 오디오 스트림들 및/또는 오디오 오브젝트들 및/또는 적응 세트들이 요청, 어드레싱 및/또는 전달될 수 있다. 특히, 다음과 같은 가능성이 존재할 수 있다:
- [0116] - 시각적 환경에 연관된 비디오 데이터 중 적어도 일부가 장면으로의 사용자의 진입 시에 서버(120)로부터 클라이언트(102)로 전달되고; 및/또는

- [0117] - 오디오 데이터(스트림들, 오브젝트들, 적응 세트들 등) 중 적어도 일부가 현재(또는 미래의) 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 및/또는 사용자의 선택/상호작용에만 기초하여 클라이언트 시스템(102)으로 전달되고; 및/또는
- [0118] - (일부 경우들에서): 일부 오디오 데이터가 (위치의 현재 또는 미래의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 및/또는 사용자의 선택에 관계없이) 현재 장면에 기초하여 클라이언트 시스템(102)으로 전달되는 반면, 나머지 오디오 데이터가 현재 또는 미래의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 및/또는 사용자의 선택에 기초하여 전달된다.
- [0119] 다양한 엘리먼트들(서버 시스템, 클라이언트 시스템, MCD 등)이 상이한 하드웨어 디바이스들에서 또는 심지어 동일한 하드웨어 디바이스들에서 엘리먼트들을 표현할 수 있음을 유의해야 한다(예를 들어, 클라이언트 및 MCD는 동일한 모바일 폰의 일부로서 구현될 수 있거나, 또는 유사하게 클라이언트는 MCD를 포함할 2차 스크린에 PC 연결될 수 있음).
- [0120] 실시예들
- [0121] 도 1a에 도시된 바와 같은 시스템(102)(클라이언트)의 일 실시예는 환경(예를 들어, 가상 환경) 내의 정의된 위치에 기초하여 (오디오) 스트림들(106)을 수신하도록 이루어지며, 그 스트림들은 비디오 및 오디오 장면(이하, 장면(150)으로 지칭됨)에 연관되는 것으로 이해될 수 있다. 일반적으로, 동일한 장면(150) 내의 상이한 위치들은 (예를 들어, 미디어 서버(120)로부터) 시스템(102)의 오디오 디코더(104)에 제공될 상이한 스트림들(106) 또는 스트림들(106)에 연관된 상이한 메타데이터를 암시한다. 시스템(102)은 미디어 소비자 디바이스(MCD)에 연결되며, 그 MCD로부터 시스템(102)은 동일한 환경에서 사용자의 위치 및/또는 가상 위치에 연관된 피드백을 수신한다. 다음에서, 환경 내의 사용자의 위치는 사용자가 즐기는 특정한 뷰포트와 연관될 수 있다(뷰포트는, 예를 들어 구 상에 투영된 직사각형 표면으로서 가정되는 의도된 표면이고, 그 표면은 사용자에게 표현됨).
- [0122] 예시적인 시나리오에서, 사용자가 VR, AR 및/또는 MR 장면(150)에서 이동할 경우, 오디오 콘텐츠는 하나 이상의 오디오 소스들(152)에 의해 가상으로 생성되는 것으로 이미징될 수 있으며, 이는 변할 수 있다. 오디오 소스들(152)은, 그들이 가상 환경 내의 위치들을 참조할 수 있다는 의미에서 가상 오디오 소스들로 이해될 수 있으며: 각각의 오디오 소스의 렌더링은 사용자 위치에 적응된다(예를 들어, 간략화된 예시에서, 오디오 소스의 레벨은, 사용자가 오디오 소스의 위치에 더 가까워질 경우 더 높고, 사용자가 오디오 소스로부터 더 멀어질 경우 더 낮음). 그럼에도 불구하고, 각각의 오디오 엘리먼트(오디오 소스)는 디코더에게 제공되는 오디오 스트림들에 인코딩된다. 오디오 스트림들은 장면 내의 다양한 위치들 및/또는 영역들에 연관될 수 있다. 예를 들어, 하나의 장면에서 가청적이지 않은 오디오 소스들(152)은 다음 장면에서, 예를 들어 VR, AR 및/또는 MR 장면(150) 내의 도어가 열릴 경우 가청적이지 않게 될 수 있다. 이어서, 사용자는 새로운 장면/환경(150)(예를 들어, 룸)에 진입하도록 선택할 수 있고, 전체 오디오 장면이 변화된다. 이러한 시나리오를 설명하려는 목적을 위해, 공간 내의 별개의 뷰포인트들이라는 용어가, 상이한 오디오 콘텐츠가 이용가능한 공간 내의(또는 VR 환경 내의) 별개의 위치로서 이용될 수 있다.
- [0123] 일반적으로, 미디어 서버(120)는 장면(150) 내의 사용자의 위치에 기초하여 특정한 장면(150)에 연관된 스트림들(106)을 제공할 수 있다. 스트림들(106)은 적어도 하나의 인코더(154)에 의해 인코딩되고 미디어 서버(120)에 제공될 수 있다. 미디어 서버(120)는 통신들(113)을 이용하여(예를 들어, 통신 네트워크를 통해) 스트림들(113)을 송신할 수 있다. 스트림들(113)의 제공은 (예를 들어, 가상 환경 내의) 사용자의 위치(110)에 기초하여 시스템(102)에 의해 착수된 요청들(112)에 기초하여 구성될 수 있다. 사용자의 위치(110)는 또한, 사용자가 즐기는 뷰포트(각각의 위치의 경우, 표현되는 하나의 단일 직사각형이 존재함) 및 (뷰포인트가 뷰포트의 중심이므로) 뷰포인트에 연관되는 것으로 이해될 수 있다. 따라서, 뷰포트의 제공은 일부 예들에서, 위치의 제공과 동일할 수 있다.
- [0124] 도 1b에 도시된 바와 같은 시스템(102)은 클라이언트 측의 다른 구성에 기초하여 (오디오) 스트림들(113)을 수신하도록 이루어진다. 인코딩 측에서의 이러한 예시적인 구현에서, 하나의 뷰포인트의 하나의 사운드 장면 부분과 연관된 각각의 이용가능한 장면(150)에 대해 하나 이상의 스트림들(106)을 생성하기 위해 사용될 수 있는 복수의 미디어 인코더들(154)이 제공된다.
- [0125] 미디어 서버(120)는, 상이한 비트레이트들로의 동일한 오디오 및 비디오 스트림들의 상이한 인코딩들을 포함하는 복수의 오디오 및 (도시되지 않은) 비디오 적응 세트들을 저장할 수 있다. 부가적으로, 미디어 서버는 모든 생성된 적응 세트들의 이용가능성을 포함할 수 있는 모든 적응 세트들의 설명 정보를 포함할 수 있다. 적응 세

트들은 또한, 하나의 적응 세트와 하나의 특정한 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보를 포함할 수 있다. 이러한 방식으로, 각각의 적응 세트는 이용가능한 오디오 장면 중 하나와 연관될 수 있다.

- [0126] 적응 세트들은, 예를 들어 완전한 오디오 장면 또는 단지 개별적인 오디오 오브젝트들을 포함할 수 있는 각각의 오디오 장면 및/또는 뷰포인트의 경계들을 설명하는 추가적인 정보를 포함할 수 있다. 하나의 오디오 장면의 경계들은, 예를 들어 구의 기하학적 좌표들(예를 들어, 중심 및 반경)로서 정의될 수 있다.
- [0127] 클라이언트 측 상의 시스템(102)은 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 관한 정보 또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보를 수신할 수 있다. 게다가, 시스템(102)은 또한, 모든 적응 세트들의 이용가능성에 관한 정보 및 하나의 적응 세트와 하나의 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보; 및/또는 (예를 들어, 완전한 오디오 장면 또는 단지 개별적인 오브젝트들을 포함할 수 있는) 각각의 오디오 장면 및/또는 뷰포인트의 "경계들"을 설명하는 정보를 수신할 수 있다. 예를 들어, 그러한 정보는 DASH 전달 환경의 경우, 미디어 프리젠테이션 설명(MPD) XML 선택스의 일부로서 제공될 수 있다.
- [0128] 시스템(102)은 콘텐츠 소비를 위해 사용되는 미디어 소비 디바이스(MCD)에 오디오 신호를 제공할 수 있다. 미디어 소비 디바이스는 또한, 위치 및 전환 데이터(110)로서 사용자 위치 및/또는 배향 및/또는 움직임의 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보) 수집을 담당한다.
- [0129] 뷰포트 프로세서(1232)는 미디어 소비 디바이스 측으로부터 상기 위치 및 전환 데이터(110)를 수신하도록 구성될 수 있다. 뷰포트 프로세서(1232)는 또한, 메타데이터에 관한 정보 및 메타데이터에서 시그널링된 ROI 및 수신단(시스템(102))에서 이용가능한 모든 정보를 수신할 수 있다. 이어서, 뷰포트 프로세서(1232)는, 수신된 및/또는 이용가능한 메타데이터로부터 수신된 및/또는 도출된 모든 정보에 기초하여, 특정한 시간 순간에 어떤 오디오 뷰포인트가 재생되어야 하는지를 판단할 수 있다. 예를 들어, 뷰포트 프로세서(1232)는, 하나의 완전한 오디오 장면이 재생될 것이거나, 하나의 새로운 오디오 장면(108)이 모든 이용가능한 오디오 장면들 중에서 생성되어야 하고, 예를 들어 복수의 오디오 장면들로부터의 일부 오디오 엘리먼트들만이 재생될 것인 반면, 이들 오디오 장면들의 다른 나머지 오디오 엘리먼트들이 재생되지 않을 것이라고 판단할 수 있다. 뷰포트 프로세서(1232)는 또한, 2개 이상의 오디오 장면들 사이의 전환이 재생되어야 하는지를 판단할 수 있다.
- [0130] 선택 부분(1230)은 뷰포트 프로세서(1232)로부터 수신된 정보에 기초하여, 수신단에 의해 수신된 정보에서 시그널링된 바와 같은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 제공될 수 있으며; 선택된 적응 세트들은 사용자의 현재 위치에서 재생되어야 하는 오디오 장면을 완전하게 설명한다. 이러한 오디오 장면은 인코딩 측에서 정의된 바와 같이 하나의 완전한 오디오 장면일 수 있거나, 또는 새로운 오디오 장면은 모든 이용가능한 오디오 장면들 중에서 생성되어야 할 수 있다.
- [0131] 부가적으로, 2개 이상의 오디오 장면들 사이의 전환이 뷰포트 프로세서(1232)의 표시에 기초하여 발생하려고 하는 경우, 선택 부분은 수신단에 의해 수신된 정보에서 시그널링된 바와 같은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 구성될 수 있으며; 선택된 적응 세트들은 가까운 미래에 재생되도록 요구될 수 있는 오디오 장면을 완전하게 설명한다(예를 들어, 사용자가 특정한 속도로 다음 오디오 장면의 방향으로 걷는다면, 다음 오디오 장면이 요구될 것이라고 예측될 수 있고, 다음 오디오 장면은 재생 전에 미리 선택됨).
- [0132] 부가적으로, 이웃한 위치들에 대응하는 일부 적응 세트들은 먼저 더 낮은 비트레이트 및/또는 더 낮은 품질 레벨로 선택될 수 있고, 예를 들어 더 낮은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택되며, 위치 변화들에 기초하여, 품질은 이들 특정 적응 세트들에 대해 더 높은 비트레이트를 선택함으로써 증가되고, 예를 들어 더 높은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택된다.
- [0133] 다운로드 및 스위칭 부분(1234)은 선택 부분으로부터 수신된 표시에 기초하여, 미디어 서버로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 요청하도록 제공될 수 있으며, 미디어 서버로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 수신하고 모든 수신된 오디오 스트림들로부터 메타데이터 정보를 추출하도록 이루어진다.
- [0134] 메타데이터 프로세서(1236)는, 수신된 오디오 스트림들에 관한 다운로드 및 스위칭 정보로부터, 수신된 각각의 오디오 스트림에 대응하는 오디오 메타데이터를 포함할 수 있는 정보를 수신하도록 제공될 수 있다. 메타데이터 프로세서(1236)는 또한, 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면을 구성하는 요구되는 오디오 엘리먼트들(152)을 선택/인에이블링시켜 모든 오디오 스트림들(113)의 단일 오디오 스트림(106)

으로의 병합을 허용하기 위해, 사용자 위치 및/또는 배향 및/또는 움직임 방향(110)에 관한 정보를 포함할 수 있는 뷰포트 프로세서(1232)로부터 수신된 정보에 기초하여, 각각의 오디오 스트림(113)과 연관된 오디오 메타데이터를 프로세싱 및 조작하도록 구성될 수 있다.

[0135] 스트림 다중화기(mux)/병합기(1238)는, 모든 수신된 오디오 스트림들(113)에 대응하는 수정된 및 프로세싱된 오디오 메타데이터를 포함할 수 있는 메타데이터 프로세서(1236)로부터 수신된 정보에 기초하여, 모든 선택된 오디오 스트림들을 하나의 오디오 스트림(106)으로 병합하도록 구성될 수 있다.

[0136] 미디어 디코더(104)는, 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면의 재생을 위해 적어도 하나의 오디오 스트림을 수신 및 디코딩하도록 이루어진다.

[0137] 다른 실시예에서, 도 1g에 도시된 바와 같은 시스템(102)은 상이한 오디오 비트레이트들 및/또는 품질 레벨들로 오디오 스트림들(106)을 수신하도록 구성될 수 있다. 이러한 실시예의 하드웨어 구성은 도 1b의 하드웨어 구성과 유사하다. 적어도 하나의 시각적 환경 장면(152)은 적어도 하나의 복수의 N개의 오디오 엘리먼트들($N \geq 2$)에 연관될 수 있으며, 각각의 오디오 엘리먼트는 시각적 환경 내의 위치 및/또는 영역에 연관된다. 적어도 하나의 복수의 N개의 오디오 엘리먼트들(152)은 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되며, 적어도 하나의 복수의 N개의 오디오 엘리먼트들(152)은 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고, 여기서 적어도 하나의 표현은 N개의 오디오 엘리먼트들(152)의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들(152) ($M < N$)을 획득하도록 N개의 오디오 엘리먼트들(152)을 프로세싱함으로써 획득된다.

[0138] N개의 오디오 엘리먼트들(152)의 프로세싱은, 예를 들어 오디오 신호들의 간단한 부가일 수 있거나, 또는 그들의 공간 위치(110)에 기초한 활성 다운믹스(downmix), 또는 오디오 신호들 사이에 위치된 새로운 가상 위치로 그들의 공간 위치를 사용하여 오디오 신호들을 렌더링하는 것일 수 있다. 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가청적인 경우 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨의 표현을 요청하도록 구성될 수 있으며, 여기서 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 덜 관련있고 그리고/또는 덜 가청적인 경우 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨의 표현을 요청하도록 이루어진다.

[0139] 도 1h는 미디어 소비 디바이스에서 재생될 비디오 스트림들(1800) 및 오디오 스트림들(106)을 수신하도록 이루어진, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 시스템(102)을 나타내는 시스템(시스템(102)일 수 있음)의 일 예를 도시하며,

[0140] 여기서, 시스템(102)은,

[0141] 사용자로의 VR, AR, MR 또는 360도 비디오 환경의 표현을 위해 비디오 스트림들(1800)로부터의 비디오 신호들(1808)을 디코딩하도록 이루어진 적어도 하나의 미디어 비디오 디코더(1804), 및

[0142] 적어도 하나의 오디오 스트림(106)으로부터의 오디오 신호들(108)을 디코딩하도록 이루어진 적어도 하나의 오디오 디코더(104)를 포함할 수 있다.

[0143] 시스템(102)은, (예를 들어, 미디어 소비 디바이스(180)로부터 피드백으로서 제공되는) 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 적어도 하나의 오디오 스트림(106) 및/또는 오디오 스트림의 하나의 오디오 엘리먼트 및/또는 하나의 적응 세트를 서버(예를 들어, 120)에 요청(112)하도록 구성될 수 있다.

[0144] 시스템(102)은 도 1a 내지 도 1g의 시스템(102)과 동일하고 그리고/또는 다음의 도 2a의 시나리오들을 획득할 수 있다.

[0145] 본 예들은 또한, 미디어 소비 디바이스[예를 들어, 재생 디바이스]에서 재생될 비디오 및/또는 오디오 스트림들을 수신하도록 이루어진, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR), 또는 360도 비디오 환경을 위한 방법을 참조하며, 그 방법은,

[0146] 사용자로의 VR, AR, MR 또는 360도 비디오 환경 장면들의 표현을 위해 비디오 스트림들로부터의 비디오 신호들을 디코딩하는 단계, 및

[0147] 오디오 스트림들로부터의 오디오 신호들을 디코딩하는 단계,

- [0148] 사용자의 현재 뷰포트 및/또는 위치 데이터 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 메타데이터 및/또는 가상 위치 데이터 및/또는 메타데이터에 기초하여 적어도 하나의 오디오 스트림을, 서버에 요청하고 그리고/또는 서버로부터 획득하는 단계를 포함한다.
- [0149] 경우 1
- [0150] 상이한 장면들/환경들(150)은 일반적으로, 서버(120)로부터의 상이한 스트림들(106)의 수신을 암시한다. 그러나, 오디오 디코더(104)에 의해 수신된 스트림들(106)은 또한 동일한 장면(150) 내의 사용자의 위치에 의해 조절될 수 있다.
- [0151] 도 2a에 도시된 제1 (시작) 순간($t=t_1$)에서, VR-환경(또는 AR-환경 또는 MR-환경)에서 제1 정의된 위치를 갖는 사용자가, 예를 들어 장면(150)에 위치된다. 데카르트(Cartesian) XYZ-시스템(예를 들어, 수평)에서, 사용자의 제1 뷰포트(위치)(110')는 좌표들(x'_u 및 y'_u)과 연관된다(여기서, 축 Z는 종이로부터 빠져나가도록 배향됨). 이러한 제1 장면(150)에서, 오디오 엘리먼트 1(152-1)에 대해 좌표들(x'_1 및 y'_1)을 그리고 오디오 엘리먼트 2(152-2)에 대해 좌표들(x'_2 및 y'_2)을 각각 갖는 2개의 오디오 엘리먼트들(152-1 및 152-2)이 위치된다. 오디오 엘리먼트 1(152-1)에 대한 사용자의 거리(d'_1)는 오디오 엘리먼트 2에 대한 사용자의 거리(d'_2)(152-1)보다 작다. 사용자 위치(뷰포트) 데이터 모두는 MCD로부터 시스템(102)으로 송신된다.
- [0152] 도 2b에 도시된 제2 예시적인 순간($t=t_2$)에서, 사용자는, 예를 들어 동일한 장면(150)에 위치되지만, 상이한 제2 위치에 위치된다. 데카르트 XY-시스템, 사용자의 제2 뷰포트(위치)(110'')는 새로운 좌표들(x'_u 및 y'_u)과 연관된다(여기서, 축 Z는 종이로부터 빠져나가도록 배향됨). 이제, 오디오 엘리먼트 1(152-1)로부터의 사용자의 거리(d'_1)는 오디오 엘리먼트 2(152-2)로부터의 사용자의 거리(d'_2)보다 크다. 사용자 위치(뷰포트) 데이터 모두는 다시 MCD로부터 시스템(102)으로 송신된다.
- [0153] 360도 환경 내에서 특정한 뷰포트를 시각화하기 위해 상기 MCD가 장착된 사용자는, 예를 들어 헤드폰들을 통해 청취하고 있을 수 있다. 사용자는 동일한 장면(150)의 도 2a 및 도 2b에 도시된 상이한 위치들에 대해 상이한 사운드들의 재생을 즐길 수 있다.
- [0154] 예를 들어, 도 2a 및 도 2b로부터의 장면 내의 임의의 위치 및/또는 임의의 전환 및/또는 뷰포트 및/또는 가상 위치 및/또는 머리 배향 및/또는 움직임 데이터는 MCD로부터 시스템(102)(클라이언트)으로 신호(110)로서 (예를 들어, 피드백에서) 주기적으로 송신될 수 있다. 클라이언트는 위치 및 전환 데이터(110' 또는 110'')(예를 들어, 뷰포트 데이터)를 서버(120)에 재송신할 수 있다. 클라이언트(102) 또는 서버(120) 중 어느 하나는 위치 및 전환 데이터(110' 또는 110'')(예를 들어, 뷰포트 데이터)에 기초하여, 어느 오디오 스트림들(106)이 현재 사용자 위치에서 정확한 오디오 장면을 재생하도록 요구되는지를 판단할 수 있다. 클라이언트는 대응하는 오디오 스트림(106)에 대한 요청(112)을 판단 및 송신할 수 있는 반면, 서버(120)는 그에 따라, 클라이언트(시스템(102))에 의해 제공된 위치 정보에 의존하여 스트림(들)(106)을 전달하도록 구성될 수 있다. 대안적으로, 서버(120)는 클라이언트(시스템(102))에 의해 제공된 위치 정보에 의존하여 스트림(들)(106)을 판단하고 그에 따라 전달할 수 있다.
- [0155] 클라이언트(시스템(102))는 장면(150)을 표현하기 위해 디코딩될 스트림들의 송신을 요청할 수 있다. 일부 예들에서, 시스템(102)은 MCD 상에서 재생될 최고의 품질 레벨에 관한 정보를 송신할 수 있다(다른 예들에서, 서버(120)는 장면 내의 사용자의 위치에 기초하여 MCD 상에서 재생될 품질 레벨을 판단함). 이에 응답하여, 서버(120)는 사용자의 위치(110' 또는 110'')에 따라 적어도 하나의 스트림(106)을 전달하기 위해, 표현될 오디오 장면과 연관된 복수의 표현들 중 하나를 선택할 수 있다. 따라서, 클라이언트(시스템(102))는 사용자의 실제(유효한) 위치(110' 또는 110'')와 연관된 사운드를 재생하기 위해 오디오 신호(108)를 사용자에게, 예를 들어 오디오 디코더(104)를 통해 전달하도록 구성될 수 있다. (적응 세트들(113)이 사용될 수 있으며: 예를 들어, 상이한 비트레이트들의 동일한 스트림들의 상이한 변형들이 사용자의 상이한 위치들에 대해 사용될 수 있음).
- [0156] (사전-프로세싱되거나 즉시 생성될 수 있는) 스트림들(106)은 클라이언트(시스템(102))에 송신될 수 있고, 특정 한 사운드 장면들과 연관된 복수의 뷰포인트들에 대해 구성될 수 있다.
- [0157] 상이한 품질들(예를 들어, 상이한 비트레이트들)이 (예를 들어, 가상) 환경 내의 사용자의 특정한 위치(예를 들어, 110' 또는 110'')에 따라 상이한 스트림들(106)에 대해 제공될 수 있음이 유의되었다. 예를 들어, 다음과 같다: 복수의 오디오 소스들(152-1 및 152-2)의 경우, 각각의 오디오 소스(152-1 및 152-2)는 장면(150) 내의

특정한 위치에 연관될 수 있다. 사용자의 위치(110' 또는 110'')가 제1 오디오 소스(152-1)에 더 가까워질수록, 제1 오디오 소스(152-2)에 연관된 스트림의 필요한 해상도 및/또는 품질이 더 높아진다. 이러한 예시적인 경우는 도 2a의 오디오 엘리먼트 1(152-1) 뿐만 아니라 도 2b의 오디오 엘리먼트 2(152-2)에 적용될 수 있다. 제2 오디오 소스(152-2)로부터 사용자의 위치(110)가 더 멀어질수록, 제2 오디오 소스(152-2)에 연관된 스트림(106)의 필요한 해상도가 더 낮아진다. 이러한 예시적인 경우는 도 2a의 오디오 엘리먼트 2(152-2) 뿐만 아니라 도 2b의 오디오 엘리먼트 1(152-1)에 적용될 수 있다.

[0158] 사실, 제1의 가까운 오디오 소스는 더 높은 레벨로 들리는 반면(그리고 그에 따라 더 높은 비트레이트로 제공되는 반면), 제2의 먼 오디오 소스는 (더 낮은 해상도를 요구하도록 허용될 수 있는) 더 낮은 레벨로 들릴 것이다.

[0159] 따라서, 클라이언트(102)에 의해 제공되는 바와 같은 환경 내의 위치(110' 또는 110'')에 기초하여, 서버(120)는 상이한 비트레이트들(또는 다른 품질)로 상이한 스트림들(106)을 제공할 수 있다. 멀리 떨어진 오디오 엘리먼트들이 높은 품질 레벨들을 요구하지 않는다는 사실에 기초하여, 그 오디오 엘리먼트들이 더 낮은 비트레이트 또는 품질 레벨로 전달되더라도, 전반적인 사용자 경험 품질이 유지된다.

[0160] 따라서, 경험 품질을 유지하면서, 상이한 사용자의 위치들에 있는 일부 오디오 엘리먼트들을 위해 상이한 품질 레벨들이 사용될 수 있다.

[0161] 이러한 솔루션이 없다면, 모든 스트림들(106)은 서버(120)에 의해 클라이언트에 최고의 비트레이트로 제공되어야 하며, 이는 서버(120)로부터 클라이언트로의 통신 채널에서 페이로드를 증가시킬 것이다.

[0162] 경우 2

[0163] 도 3(경우 2)은 다른 예시적인 시나리오(공간 XYZ의 수직 평면 XZ에서 표현됨, 여기서 축 Y는 종으로 진입하는 것으로 표현됨)를 갖는 일 실시예를 도시하고, 여기서 사용자는 제1 VR, AR 및/또는 MR 장면 A(150A)에서 이동하고, 도어를 열고 걸어 들어가며(전환(150AB)), 이는, 시간(t_1)의 제1 장면(150A)으로부터 시간(t_2)의 전환 위치(150AB)를 걸친 시간(t_3)의 다음(제2) 장면 B(150B)으로의 오디오의 전환을 암시한다.

[0164] 시간(t_1)에서, 사용자는 제1 VR, AR 및/또는 MR 장면의 x-방향에서 위치(x_1)에 있을 수 있다. 시간(t_3)에서, 사용자는 위치(x_3)에서 상이한 제2 VR, AR 및/또는 MR 장면 B(150B)에 있을 수 있다. 순간(t_2)에서, 사용자가 도어(예를 들어, 가상 도어)를 열고 걸어 들어가는 동안, 사용자는 전환 위치(150AB)에 있을 수 있다. 따라서, 전환은 제1 장면(150A)으로부터 제2 장면(150B)으로의 오디오 정보의 전환을 암시한다.

[0165] 이러한 맥락에서, 사용자는, 예를 들어 (도 1a에 도시된 바와 같이 제1 뷰포인트(A)에 의해 특징지어지는) 제1 VR-환경으로부터 (도 1a에 도시된 바와 같이 제2 뷰포인트(B)에 의해 특징지어지는) 제2 VR-환경으로 자신의 위치(110)를 변화시킨다. 특정한 경우에서, 예를 들어 x-방향에서 위치(x_2)에 위치된 도어를 통한 전환 동안, 일부 오디오 엘리먼트들(152A 및 152B)은 뷰포인트들 둘 모두(위치들 A 및 B)에 존재할 수 있다.

[0166] (MCD가 장착된) 사용자는 도어를 향해 자신의 위치(110)를 변화시키며(x_1 - x_3), 이는 전환 위치(x_2)에서, 오디오 엘리먼트들이 제1 장면(150A) 및 제2 장면(150B) 둘 모두에 속한다는 것을 암시할 수 있다. MCD는 새로운 위치 및 전환 데이터(110)를 클라이언트에 송신하고, 클라이언트는 이를 미디어 서버(120)에 재송신한다. 사용자는, 제1 및 제2 위치들(x_1 및 x_3) 사이의 중간 위치(x_2)에 의해 정의된 적절한 오디오 소스들을 청취할 수 있게 될 수 있다.

[0167] 제1 위치(x_1)로부터 제2 위치(x_3)까지의 임의의 전환 및 임의의 위치가 이제 MCD로부터 클라이언트에 주기적으로 (예를 들어, 연속적으로) 송신된다. 클라이언트(102)는 위치 및 전환 데이터(110)(x_1 - x_3)를 미디어 서버(120)에 재송신할 수 있으며, 미디어 서버(120)는 그에 따라, 수신된 위치 및 전환 데이터(110)(x_1 - x_3)에 의존하여, 현실화된 적응 세트(113')의 형태로, 예를 들어, 사전프로세싱된 스트림들(106)의 새로운 세트의 하나의 전용 아이템을 전달하도록 이루어진다.

[0168] 미디어 서버(120)는, 최고의 비트레이트를 디스플레이하기 위한 MCD의 능력 뿐만 아니라 하나의 위치로부터 다른 위치로의 사용자의 움직임 동안 사용자의 위치 및 전환 데이터(110)(x_1 - x_3)에 관해, 전송된 정보와 연관된 복수의 표현들 중 하나를 선택할 수 있다. (이러한 맥락에서, 적응 세트들을 사용하는 것이 가능하며: 미디어 서

버(120)는 MCD의 렌더링 용량을 방해하지 않으면서, 어느 적응 세트(113')가 사용자의 가상 전환을 최적으로 표현하는지를 판단할 수 있음).

[0169] 따라서, 미디어 서버(120)는 위치들의 전환에 따라 (예를 들어, 새로운 적응 세트(113')로서) 전용 스트림(106)을 전달할 수 있다. 그에 따라, 클라이언트(102)는, 예를 들어 미디어 오디오 디코더(104)를 통해 오디오 신호(108)를 사용자(140)에게 전달하도록 구성될 수 있다.

[0170] (즉시 생성되고 그리고/또는 사전프로세싱된) 스트림들(106)은 주기적으로(예를 들어, 연속적으로) 현실화된 적응 세트(113')로 클라이언트(102)에 송신될 수 있다.

[0171] 사용자가 도어를 통해 걸을 경우, 서버(120)는 제1 장면(150A)의 스트림들(106) 및 제2 장면(150B)의 스트림들(106) 둘 모두를 송신할 수 있다. 이것은, 사용자에게 실제 인상을 제공하기 위해 이들 스트림들(106)을 동시에 믹싱하거나 다중화하거나 조립하거나 재생하기 위한 것이다. 따라서, 사용자의 위치(110)(예를 들어, "도어에 대응하는 위치")에 기초하여, 서버(120)는 상이한 스트림들(106)을 클라이언트에 송신한다.

[0172] 이러한 경우에도, 상이한 스트림들(106)이 동시에 들릴 것이므로, 그들은 상이한 해상도들을 가질 수 있고, 상이한 해상도들로 서버(120)로부터 클라이언트에 송신될 수 있다. 사용자가 전환을 완료하고 제2(위치) 장면(150A)에 있을 경우(그리고 그의 뒤에서 도어를 닫을 경우), 서버(120)가 제1 장면(150)의 스트림들(106)을 송신하는 것을 감소시키거나 억제할 가능성이 존재할 것이다(서버(120)가 클라이언트(102)에게 스트림들을 이미 제공했던 경우, 클라이언트(102)는 그들을 사용하지 않기로 판단할 수 있음).

[0173] 경우 3

[0174] 도 4(경우 3)는 다른 예시적인 시나리오(공간 XYZ의 수직 평면 XZ에서 표현됨, 여기서 축 Y는 종이로 진입하는 것으로 표현됨)를 갖는 일 실시예를 도시하며, 여기서 사용자는 VR, AR 및/또는 MR 장면 A(150A)에서 이동하며, 이는 시간(t_1)의 제1 위치로부터 시간(t_2)의 또한 제1 장면(150A) 내의 제2 위치로의 오디오의 전환을 암시한다. 제1 위치의 사용자는 벽으로부터의 거리(d_1)로 시간(t_1)에서 벽으로부터 멀리 있을 수 있으며; 벽으로부터의 거리(d_2)로 시간(t_2)에서 벽에 가까워질 수 있다. 여기서, $d_1 > d_2$ 이다. 거리(d_1)에서, 사용자가 장면(150A)의 소스(152A)만을 들지만, 사용자는 또한 벽을 넘어 장면(150B)의 소스(152B)를 들을 수 있다.

[0175] 사용자가 제2 위치(d_2)에 있을 경우, 클라이언트(102)는 사용자의 위치(110)(d_2)에 관한 데이터를 서버(120)에 전송하고, 제1 장면(150A)의 오디오 스트림들(106) 뿐만 아니라 제2 장면(150B)의 오디오 스트림들(106)을 서버(120)로부터 수신한다. 예를 들어, 서버(120)에 의해 제공된 메타데이터에 기초하여, 클라이언트(102)는 낮은 볼륨으로의 (벽을 넘은) 제2 장면(150B)의 스트림들(106)의, 예를 들어 디코더(104)를 통한 재생을 야기할 것이다.

[0176] 이러한 경우에도, 제2 장면(150B)의 스트림들(106)의 비트레이트(품질)는 낮을 수 있으며, 따라서 서버(120)로부터 클라이언트로의 감소된 송신 페이로드를 요구한다. 특히, 클라이언트(및/또는 뷰포트)의 위치(110)(d_1 , d_2)는 서버(120)에 의해 제공되는 오디오 스트림들(106)을 정의한다.

[0177] 예를 들어, 시스템(102)은 제1 현재 환경에 연관된 제1 현재 장면(150A)에 연관된 스트림들을 획득하도록 구성될 수 있으며, 장면의 경계(예를 들어, 벽에 대응함)로부터의 사용자의 위치 또는 가상 위치의 거리가 미리 결정된 임계치 미만일 경우(예를 들어, $d_2 < d_{\text{threshold}}$ 일 경우), 시스템(102)은 제2 장면(150B)에 연관된 제2의 인접한 및/또는 이웃한 환경에 연관된 오디오 스트림들을 추가로 획득한다.

[0178] 경우 4

[0179] 도 5a 및 도 5b는 다른 예시적인 시나리오(공간 XYZ의 수평 평면 XZ에서 표현됨, 여기서 축 Z는 종이로부터 빠져나가는 것으로 표현됨)를 갖는 일 실시예를 도시하며, 여기서 사용자는 하나의 동일한 VR, AR 및/또는 MR 장면(150)에 위치되지만, 예를 들어 2개의 오디오 엘리먼트들에 대해 상이한 거리들로 상이한 순간들에 위치된다.

[0180] 도 5a에 도시된 제1 순간($t=t_1$)에서, 사용자는, 예를 들어 제1 위치에 위치된다. 이러한 제1 위치에서, 제1 오디오 엘리먼트 1(152-1) 및 제2 오디오 엘리먼트 2(152-2)는 MCD가 장착된 사용자로부터 각각 거리들(d_1 및 d_2)에 (예를 들어, 가상으로) 위치된다. 이러한 경우, 거리들(d_1 및 d_2)은 정의된 임계 거리($d_{\text{threshold}}$)보다 클 수 있으며, 따라서 시스템(102)은 오디오 엘리먼트들 둘 모두를 하나의 단일 가상 소스(152-3)로 그룹화하도록 이

루어진다. 단일 가상 소스의 위치 및 속성들(예를 들어, 공간 범위)은, 단일 가상 소스가 2개의 소스들에 의해 생성된 본래의 사운드 필드를 가능한 양호하게 모방하는 그러한 방식으로, 예를 들어 본래의 2개의 소스들의 위치들에 기초하여 계산될 수 있다(예를 들어, 2개의 매우 양호하게 로컬화된 포인트 소스들은 그들 사이의 거리의 중간에서 단일 소스로서 재생될 수 있음). 사용자 위치 데이터(110)(d_1 , d_2)는 MCD로부터 시스템(102)(클라이언트)에 그리고 후속하여 서버(120)에 송신될 수 있으며, 서버(120)는 서버 시스템(120)에 의해 렌더링될 적절한 오디오 스트림(106)을 전송하기로 판단할 수 있다(다른 실시예들에서, 어느 스트림들이 서버(120)로부터 송신될지를 판단하는 것은 클라이언트(102)임). 오디오 엘리먼트들 둘 모두를 하나의 단일 가상 소스(152-3)로 그룹화함으로써, 서버(120)는 전송된 정보와 연관된 복수의 표현들 중 하나를 선택할 수 있다. (예를 들어, 예컨대 하나의 단일 채널과 그에 따라 연관된 적응 세트(113')와 함께 전용 스트림(106)을 그에 따라 전달하는 것이 가능함). 결과적으로, 사용자는 실제 오디오 엘리먼트들 1(152-1) 및 2(152-2) 사이에 위치한 단일 가상 오디오 엘리먼트(152-3)로부터 송신되는 것으로서 오디오 신호를 MCD를 통해 수신할 수 있다.

[0181] 도 5b에 도시된 제2 순간($t=t_2$)에서, 도 5a에서와 같이 동일한 VR-환경에서 제2의 정의된 위치를 갖는 사용자는, 예를 들어 동일한 장면(150)에 위치된다. 이러한 제2 위치에서, 2개의 오디오 엘리먼트들(152-1 및 152-2)은 사용자로부터 각각 거리들(d_3 및 d_4)에 (예를 들어, 가상으로) 위치된다. 거리들(d_3 및 d_4) 둘 모두가 임계 거리($d_{\text{threshold}}$)만큼 작을 수 있으며, 따라서 오디오 엘리먼트들(152-1 및 152-2)의 하나의 단일 가상 소스(152-3)로의 그룹화는 더 이상 사용되지 않는다. 사용자 위치 데이터는 MCD로부터 시스템(102)에 그리고 후속하여 서버(120)에 송신되며, 서버(120)는 시스템 서버(120)에 의해 렌더링될 다른 적절한 오디오 스트림(106)을 전송하기로 판단할 수 있다(다른 실시예들에서, 이러한 판단은 클라이언트(102)에 의해 행해짐). 오디오 엘리먼트들을 그룹화하는 것을 회피함으로써, 서버(120)는 전송된 정보와 연관된 상이한 표현을 선택하여, 각각의 오디오 엘리먼트에 대해 상이한 채널들과 그에 따라 연관된 적응 세트(113')와 함께 전용 스트림(106)을 그에 따라 전달할 수 있다. 결과적으로, 사용자는 2개의 상이한 오디오 엘리먼트들 1(152-1) 및 2(152-2)로부터 송신되는 것으로서 오디오 신호(108)를 MCD를 통해 수신할 수 있다. 따라서, 사용자의 위치(110)가 오디오 소스들 1(152-1) 및 2(152-2)에 더 가까워질수록, 오디오 소스들에 연관된 스트림의 필요한 품질 레벨이 더 높게 선택되어야 한다.

[0182] 실제로, 도 5b에 도시된 바와 같이, 오디오 소스들 1(152-1) 및 2(152-2)이 사용자에게 더 가깝게 위치될수록, 레벨이 더 높게 조정되어야 하며, 따라서 오디오 신호들(108)은 더 높은 품질 레벨로 렌더링될 수 있다. 반대로, 도 5b에 표현된 원격 위치된 오디오 소스들 1 및 2는 단일 가상 소스에 의해 재생되는 바와 같이 더 낮은 레벨로 들려야 하며, 따라서 예를 들어 더 낮은 품질 레벨로 렌더링된다.

[0183] 유사한 구성에서, 복수의 오디오 엘리먼트들이 사용자의 앞에 위치될 수 있으며, 이들 모두는 사용자로부터 임계 거리보다 큰 거리들에 위치된다. 일 실시예에서, 5개의 오디오 엘리먼트들의 2개의 그룹들 각각은 2개의 가상 소스들로 결합될 수 있다. 사용자 위치 데이터는 MCD로부터 시스템(102)에 그리고 후속하여 서버(120)에 송신되며, 서버(120)는 시스템 서버(120)에 의해 렌더링될 적절한 오디오 스트림(106)을 전송하기로 판단할 수 있다. 10개의 오디오 엘리먼트들 모두를 단지 2개의 단일 가상 소스들로만 그룹화함으로써, 서버(120)는 전송된 정보와 연관된 복수의 표현들 중 하나를 선택하여, 예를 들어 2개의 단일 오디오 엘리먼트들과 그에 따라 연관된 적응 세트(113')와 함께 전용 스트림(106)을 그에 따라 전달할 수 있다. 결과적으로, 사용자는 실제 오디오 엘리먼트들과 동일한 위치 영역에 위치된 2개의 별개의 가상 오디오 엘리먼트들로부터 송신되는 것으로서 오디오 신호를 MCD를 통해 수신할 수 있다.

[0184] 후속 시간 순간에, 사용자는 복수(10개)의 오디오 엘리먼트들에 접근하고 있다. 이러한 후속 장면에서, 오디오 엘리먼트들 모두는 임계 거리($d_{\text{threshold}}$)만큼 작은 거리들에 위치되며, 따라서 시스템(102)은 오디오 엘리먼트들의 그룹화를 종료하도록 이루어진다. 새로운 사용자 위치 데이터는 MCD로부터 시스템(102)에 그리고 후속하여 서버(120)에 송신되며, 서버(120)는 서버 시스템(120)에 의해 렌더링될 다른 적절한 오디오 스트림(106)을 전송하기로 판단할 수 있다. 오디오 엘리먼트들을 그룹하지 않음으로써, 서버(120)는 전송된 정보와 연관된 상이한 표현을 선택하여, 각각의 오디오 엘리먼트에 대해 상이한 채널들과 그에 따라 연관된 적응 세트(113')와 함께 전용 스트림(106)을 그에 따라 전달할 수 있다. 결과적으로, 사용자는 10개의 상이한 오디오 엘리먼트들로부터 송신되는 것으로서 오디오 신호를 MCD를 통해 수신할 수 있다. 따라서, 사용자의 위치(110)가 오디오 소스들에 더 가까워질수록, 오디오 소스들에 연관된 스트림의 필요한 해상도가 더 높게 선택되어야 한다.

[0185] 경우 5

- [0186] 도 6(경우 5)은, 예시적인 3개의 상이한 방향들(각각은 상이한 뷰포트(160-1, 160-2, 160-3)에 연관됨)로 지향될 수 있는 미디어 소비자 디바이스(MCD)를 착용한, 하나의 단일 장면(150)의 일 위치에 위치한 사용자(140)를 도시한다. 도 6에 도시된 바와 같은 이들 방향들은, 예를 들어 도 6의 하단 부분에 180°로 위치한 제1 뷰포인트(801), 예를 들어 도 6의 우측 상에 90°로 위치한 제2 뷰포인트(802), 및 예를 들어 도 6의 상부 부분에 0°로 위치한 제3 뷰포인트(803)를 가리키는 극 좌표계 및/또는 데카르트 XY-시스템에서의 배향(예를 들어, 각도 배향)을 가질 수 있다. 이들 뷰포인트들 각각은 미디어 소비자 디바이스(MCD)를 착용한 사용자(140)의 배향에 연관되며, 사용자는 MCD의 배향에 따라 대응하는 오디오 신호(108)를 렌더링하는 MCD에 의해 디스플레이된 특정 뷰포트를 제공받는 중심에 위치된다.
- [0187] 이러한 특정한 VR 환경에서, 제1 오디오 엘리먼트 s1(152)은, 예를 들어 180°로 위치한 뷰포인트의 이웃에서 제1 뷰포트(160-1)에 위치되고, 제2 오디오 엘리먼트 s2(152)는, 예를 들어, 180°로 위치한 뷰포인트의 이웃에서 제3 뷰포트(160-3)에 위치된다. 자신의 배향을 변화시키기 전에, 사용자(140)는 뷰포인트(801)(뷰포트(160-1))을 향한 제1 배향에서, 자신의 실제(유효) 위치와 연관된 사운드가 오디오 엘리먼트 s2보다 오디오 엘리먼트 s1으로부터의 소리가 더 크다는 것을 경험한다.
- [0188] 자신의 배향을 변화시킴으로써, 사용자(140)는 뷰포인트(802)를 향한 제2 배향에서, 자신의 실제 위치(110)와 연관된 사운드가 오디오 엘리먼트들 s1 및 s2 둘 모두로부터의 옆에서 오는 거의 동일한 라우드니스(loudness)이라는 것을 경험할 수 있다.
- [0189] 마지막으로, 자신의 배향을 변화시킴으로써, 사용자(140)는 뷰포인트(801)(뷰포트(160-3))를 향한 제3 배향에서, 오디오 엘리먼트 2와 연관된 사운드가 오디오 엘리먼트 s1에 연관된 사운드보다 소리가 크다는 것을 경험할 수 있다(사실, 오디오 엘리먼트 2로부터 나오는 사운드는 전방으로부터 도달하는 반면, 오디오 엘리먼트 1로부터 나오는 사운드는 후방으로부터 도달함).
- [0190] 따라서, 상이한 뷰포트들 및/또는 배향들 및/또는 가상 위치 데이터는 상이한 비트레이트들 및/또는 품질들에 연관될 수 있다.
- [0191] 다른 경우들 및 실시예들
- [0192] 도 7a는 다이어그램에서 동작 단계들의 시퀀스의 형태로 시스템에 의해 오디오 스트림들을 수신하기 위한 방법의 일 실시예를 도시한다. 임의의 순간에, 시스템(102)의 사용자는 자신의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치와 연관된다. 특정한 순간에, 시스템은 도 7a의 단계(701)에서, 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 기초하여, 재생될 오디오 엘리먼트들을 결정할 수 있다. 따라서, 다음 단계(703)에서, 각각의 오디오 엘리먼트에 대한 관련성 및 가청성 레벨이 결정될 수 있다. 위의 도 6에서 설명된 바와 같이, VR 환경은 사용자의 이웃에 또는 멀리 떨어진 특정한 장면(150)에 위치한 상이한 오디오 엘리먼트들을 가질 수 있지만, 또한 360도 주변의 특정 배향을 가질 수 있다. 이들 인자들 모두는 상기 오디오 엘리먼트 각각에 대한 관련성 및 가청성 레벨을 결정한다.
- [0193] 다음 단계(705)에서, 시스템(102)은 미디어 서버(120)로부터 오디오 엘리먼트들 각각에 대한 결정된 관련성 및 가청성 레벨에 따라 오디오 스트림들을 요청할 수 있다.
- [0194] 다음 단계(707)에서, 시스템(102)은 미디어 서버(120)에 의해 그에 따라 준비된 오디오 스트림들(113)을 수신할 수 있으며, 여기서 상이한 비트레이트들을 갖는 스트림들은 전술한 단계들에서 결정된 바와 같은 관련성 및 가청성 레벨을 반영할 수 있다.
- [0195] 다음 단계(709)에서, 시스템(102)(예를 들어, 오디오 디코더)은 수신된 오디오 스트림들(113)을 디코딩할 수 있어서, 단계(711)에서, 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 따라 특정한 장면(150)이(예를 들어, MCD에 의해) 재생된다.
- [0196] 도 7b는 앞서 설명된 동작 다이어그램의 시퀀스에 따른 미디어 서버(120)와 시스템(102) 사이의 상호작용을 도시한다. 특정한 순간에, 미디어 서버는 전술한 장면(150)의 관련있는 오디오 엘리먼트들의 앞서 결정된 더 낮은 관련성 및 가청성 레벨에 따라, 더 낮은 비트레이트로 오디오 스트림(750)을 송신할 수 있다. 시스템은 후속 순간(752)에서, 상호작용 또는 위치 데이터의 변화가 발생한다고 결정할 수 있다. 그러한 상호작용은, 예를 들어 동일한 장면(150)에서의 위치 데이터의 변화 또는, 예를 들어 사용자가 도어 핸들에 의해 제공되는 도어에 의해 제1 장면으로부터 분리된 제2 장면으로 진입하려고 시도하는 동안 도어 핸들을 활성화시키는 것으로부터 초래될

수 있다.

- [0197] 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치의 변화는 시스템(102)에 의해 미디어 서버(120)에 전송되는 요청(754)을 초래할 수 있다. 이러한 요청은 후속 장면(150)에 대해 결정된 관련있는 오디오 엘리먼트들의 더 높은 관련성 및 가청성 레벨을 반영할 수 있다. 요청(754)에 대한 응답으로서, 미디어 서버는 더 높은 비트레이트로 스트림(756)을 송신할 수 있으며, 이는 임의의 현재 사용자의 가상 위치에서의 시스템(102)에 의한 장면(150)의 그럴듯하고 현실적인 재생을 가능하게 한다.
- [0198] 도 8a는, 또한 다이어그램에서 동작 단계들의 시퀀스의 형태로 시스템에 의해 오디오 스트림들을 수신하기 위한 방법의 다른 실시예를 도시한다. 특정한 순간(801)에, 제1 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치의 결정이 수행될 수 있다. 긍정적인 경우를 추론함으로써, 단계(803)에서, 낮은 비트레이트에 의해 정의된 제1 위치에 연관된 스트림들의 요청이 준비되고 시스템(102)에 의해 송신될 수 있다.
- [0199] 3개의 상이한 결과들을 갖는 결정 단계(805)가 후속 순간에 수행될 수 있다. 예를 들어, 후속 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치에 관한 예측적인 판단을 결정하기 위해 하나 또는 2개의 정의된 임계치(들)가 이러한 단계에서 관련있을 수 있다. 따라서, 제1 및/또는 제2 임계치와의 비교가 제2 위치로의 변화 확률에 관해 수행되어, 예를 들어 수행될 3개의 상이한 후속 단계들을 초래할 수 있다.
- [0200] 예를 들어, 매우 낮은 확률(예를 들어, 제1 미리 결정된 임계치와의 위의 비교에 연관됨)을 반영하는 결과에서, 새로운 비교 단계(801)가 수행될 것이다.
- [0201] 낮은 확률(예를 들어, 제1 미리 결정된 임계치보다 높지만, 예들에서는 제1 임계치보다 높은 제2 미리 결정된 임계치보다 낮음)을 반영하는 결과에서, 단계(809)에서 낮은 비트레이트의 오디오 스트림들(113)에 대한 요청이 초래될 수 있다.
- [0202] 높은 확률(예를 들어, 제2 미리 결정된 임계치보다 높음)을 반영한 결과에서, 단계(807)에서 높은 비트레이트의 오디오 스트림들(113)에 대한 요청이 수행될 수 있다. 따라서, 단계들(807 또는 809)을 수행한 이후 수행될 후속 단계는 다시 결정 단계(801)일 수 있다.
- [0203] 도 8b는 앞서 설명된 동작 다이어그램의 시퀀스 중 오직 하나에 따른 미디어 서버(120)와 시스템(102) 사이의 상호작용을 도시한다. 특정한 순간에, 미디어 서버는 전송한 장면(150)의 오디오 엘리먼트들의 앞서 결정된 낮은 관련성 및 가청성 레벨에 따라, 낮은 비트레이트로 오디오 스트림(850)을 송신할 수 있다. 시스템은 후속 순간(852)에서, 상호작용이 예측적으로 발생할 것이라고 결정할 수 있다. 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치의 예측적인 변화는 시스템(102)에 의해 미디어 서버(120)에 전송되는 적절한 요청(854)을 초래할 수 있다. 이러한 요청은 개개의 후속 장면(150)에 대해 요구되는 바와 같이 오디오 엘리먼트들의 가청성 레벨에 따라 높은 비트레이트와 연관된 제2 위치에 도달할 높은 확률에 관해, 위에서 설명된 경우들 중 하나를 반영할 수 있다. 응답으로서, 미디어 서버는 더 높은 비트레이트로 스트림(856)을 송신할 수 있으며, 이는 임의의 현재 사용자의 가상 위치에서의 시스템(102)에 의한 장면(150)의 그럴듯하고 현실적인 재생을 가능하게 한다.
- [0204] 도 1c에 도시된 바와 같은 시스템(102)은 클라이언트 측의 다른 구성에 기초하여 오디오 스트림들(113)을 수신하도록 이루어지며, 여기서 시스템 아키텍처는 복수의 오디오 디코더들(1320, 1322)을 사용하는 솔루션에 기초하여 별개의 뷰포인트들을 사용할 수 있다. 클라이언트 측에서, 예를 들어, 시스템(102)은, 예를 들어 복수의 오디오 엘리먼트들이 비활성화되어 있는 메타데이터 프로세서(1236)에 의해 표시된 바와 같이 개별 오디오 스트림들을 디코딩하도록 구성될 수 있는 복수의 오디오 디코더들(1320, 1322)을 부가적으로 또는 대안적으로 포함하는 도 1b에 설명된 시스템의 일부들을 구현할 수 있다.
- [0205] 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 최종 오디오 장면을 재생하도록 이루어진 믹서/렌더러(1238)가 시스템(102)에서 제공될 수 있으며, 즉 예를 들어, 그 특정 위치에서 가청적이지 않은 오디오 엘리먼트들 중 일부는 디스에이블링되거나 렌더링되지 않아야 한다.
- [0206] 도 1d, 도 1e 및 도 1f에 도시된 다음의 실시예들은 유연한 적응 세트들을 갖는 별개의 뷰포인트들에 대한 독립적인 적응 세트들에 기초한다. 사용자가 VR 환경에서 이동하는 경우, 오디오 장면은 연속적인 방식으로 변화될 수 있다. 양호한 오디오 경험을 보장하기 위해, 특정한 시간 순간에 오디오 장면을 조립하는 모든 오디오 엘리먼트들은, 최종 오디오 장면을 생성하기 위해 위치 정보를 이용할 수 있는 미디어 디코더에 이용가능하게 되어

야 할 수 있다.

- [0207] 복수의 미리-정의된 위치들에 대해 콘텐츠가 미리-인코딩되면, 시스템은, 오디오 장면들이 중첩되지 않고 사용자가 하나의 위치로부터 다음 위치로 "점프/스위칭"할 수 있다는 가정 하에서, 이들 특정 위치들에서 이들 오디오 장면들의 정확한 재생을 제공할 수 있다.
- [0208] 그러나, 사용자가 하나의 위치로부터 다음 위치로 "걷는" 경우들에서, 2개(또는 그 이상)의 오디오 장면들로부터의 오디오 엘리먼트들이 동시에 가청적일 수 있다. 이러한 사용 경우들에 대한 솔루션은 이전의 시스템 예들에서 제공되었으며, 여기서 (단일 미디어 디코더 또는 복수의 미디어 디코더를 갖는 다중화기 및 부가적인 믹서/렌더러를 사용하여) 복수의 오디오 스트림들을 디코딩하기 위해 제공되는 메커니즘들과 독립적으로, 완전한 오디오 장면들을 설명하는 오디오 스트림들은 클라이언트에 제공되어야 한다.
- [0209] 다음에서, 복수의 오디오 스트림들 사이에 공통 오디오 엘리먼트들의 개념을 도입함으로써 최적화가 제공된다.
- [0210] 양상들 및 실시예들에 대한 논의
- [0211] 솔루션 1: 별개의 위치들(뷰포인트들)에 대한 독립적인 적응 세트들.
- [0212] 설명된 문제를 해결하는 하나의 방식은 각각의 위치에 대해 완전히 독립적인 적응 세트들을 사용하는 것이다. 솔루션의 더 양호한 이해를 위해, 도 1a이 예시적인 시나리오로서 사용된다. 이러한 예에서, 사용자가 이동할 수 있어야 하는 완전한 VR 호나경을 생성하기 위해 (3개의 상이한 오디오 장면들을 포함하는) 3개의 상이한 별개의 뷰포인트들이 사용된다. 따라서:
- [0213] ● 수 개의 독립적이거나 중첩하는 오디오 장면들이 복수의 오디오 스트림들로 인코딩된다. 각각의 오디오 장면들에 대해, 하나의 메인 스트림이 사용 경우, 하나의 메인 스트림 및 부가적인 보조 스트림들에 의존하여 사용될 수 있다(예를 들어, 상이한 언어들을 포함하는 일부 오디오 오브젝트들은 효율적인 전달을 위해 독립적인 스트림들로 인코딩될 수 있음). 제공된 예에서, 오디오 장면 A는 2개의 스트림들(A1 및 A2)로 인코딩되고, 오디오 장면 B는 3개의 스트림들(B1, B2, 및 B3)로 인코딩되는 반면, 오디오 장면 C는 3개의 스트림들(C1, C2 및 C3)로 인코딩된다. 오디오 장면 A 및 오디오 장면 B가 복수의 공통 엘리먼트들(이러한 예에서는 2개의 오디오 오브젝트들)을 공유함을 유의해야 한다. 모든 장면이 (예를 들어, 비-VR 재생 디바이스들 상에서의 독립적인 재생을 위해) 완전하고 독립적이어야 하므로, 공통 엘리먼트들은 각각의 장면들에 대해 2회 인코딩되어야 한다.
- [0214] ● 모든 오디오 스트림들은 상이한 비트레이트들(즉, 상이한 표현들)로 인코딩되어, 네트워크 연결에 의존하여 효율적인 비트레이트 적응을 허용한다(즉, 고속 연결을 사용하는 사용자들에 대해서는 높은 비트레이트 코딩된 버전이 전달되는 반면, 더 낮은 속도의 네트워크 연결을 갖는 사용자들에 대해서는 더 낮은 비트레이트 버전이 전달됨).
- [0215] ● 오디오 스트림들은 미디어 서버 상에 저장되며, 여기서 각각의 오디오 스트림에 대해, 상이한 비트레이트들(즉, 상이한 표현들)의 상이한 인코딩들은 하나의 적응 세트로 그룹화되는 데, 적절한 데이터가 모든 생성된 적응 세트들의 이용가능성을 시그널링한다.
- [0216] ● 적응 세트들에 부가적으로, 미디어 서버는 각각의 오디오 장면의 위치 "경계들" 및 (예를 들어 완전한 오디오 장면 또는 단지 개별적인 오브젝트들을 포함할 수 있는) 각각의 적응 세트에 대한 그의 관계에 관한 정보를 수신한다. 이러한 방식으로, 각각의 적응 세트는 이용가능한 오디오 장면 중 하나와 연관될 수 있다. 하나의 오디오 장면의 경계들은, 예를 들어 구의 기하학적 좌표들(예를 들어, 중심 및 반경)로서 정의될 수 있다.
- [0217] ○ 각각의 적응 세트는, 사운드 장면 또는 오디오 엘리먼트들이 활성화된 위치들에 관한 설명 정보를 또한 포함한다. 예를 들어, 하나의 보조 스트림이 하나 또는 수 개의 오브젝트들을 포함하면, 적응 세트는, 오브젝트들이 가청적인 위치들(예를 들어, 구의 중심의 좌표들 및 반경)과 같은 정보를 포함할 수 있다.
- [0218] ● 미디어 서버는 각각의 적응 세트와 연관된 위치 "경계들"에 관한 정보를 클라이언트, 예를 들어 DASH 클라이언트에 제공한다. 예를 들어, 이것은 DASH 전달 환경의 경우, 미디어 프리젠테이션 설명(MPD) XML 신텍스로 임베딩될 수 있다.
- [0219] ● 클라이언트는 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보)를 수신한다.
- [0220] ● 클라이언트는 각각의 적응 세트에 관한 정보를 수신하고, 그 정보 및 사용자 위치 및/또는 배향 및/또는 움직임 방향(또는, 예를 들어 x,y,z 좌표들 및/또는 요, 피치, 롤 값들을 포함하는, 사용자 액션들에 의해 트리거

링된 변화들을 특성화하는 임의의 정보)에 기초하여, 클라이언트는 사용자의 현재 위치에서 재생되어야 하는 오디오 장면을 완전하게 설명하는 하나 이상의 적응 세트들을 선택한다.

- [0221] ● 클라이언트는 하나 이상의 적응 세트들을 요청한다.
- [0222] ○ 게다가, 클라이언트는 하나 초과 오디오 장면을 완전하게 설명하는 더 많은 적응 세트들을 선택하고, 하나 초과 오디오 장면에 대응하는 오디오 스트림들을 사용하여, 사용자의 현재 위치에서 재생되어야 하는 새로운 오디오 장면을 생성할 수 있다. 예를 들어, 사용자가 VR 환경에서 걷고 있고 일 시간 순간에 그 사이에(또는 2 개의 오디오 장면들이 가청적 효과들을 갖는 장소에 위치된 위치에) 위치되는 경우.
- [0223] ○ 일단 오디오 스트림들이 이용가능하면, 복수의 미디어 디코더들은 개별 오디오 스트림들을 디코딩하는 데 사용될 수 있고, 부가적인 믹서/렌더러(1238)는 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 최종 오디오 장면을 재생하는 데 사용될 수 있다(즉 예를 들어, 그 특정 위치에서 가청적이지 않은 오디오 엘리먼트들 중 일부는 디스에이블링되거나 렌더링되지 않아야 함).
- [0224] ○ 대안적으로, 메타데이터 프로세서(1236)는 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 모든 오디오 스트림들과 연관된 오디오 메타데이터를 조작하는 데 사용되어:
- [0225] ■ 새로운 오디오 장면을 구성하는 요구되는 오디오 엘리먼트들(152)을 선택/인에이블링하고;
- [0226] ■ 그리고 모든 오디오 스트림들의 단일 오디오 스트림으로의 병합을 허용할 수 있다.
- [0227] ● 미디어 서버는 요구되는 적응 세트들을 전달한다.
- [0228] ● 대안적으로, 클라이언트는 사용자 위치설정에 관한 정보를 미디어 서버에 제공하고, 미디어 서버는 요구되는 적응 세트들에 관한 표시를 제공한다.
- [0229] 도 1b는 다음을 포함하는 그러한 시스템의 다른 예시적인 구현을 도시한다:
- [0230] ● 인코딩 측에서,
- [0231] ○ 하나의 뷰포인트의 하나의 사운드 장면 부분과 연관된 각각의 이용가능한 오디오 장면에 대해 하나 이상의 오디오 스트림들을 생성하기 위해 사용될 수 있는 복수의 미디어 인코더들,
- [0232] ○ 하나의 뷰포인트의 하나의 비디오 장면 부분과 연관된 각각의 이용가능한 비디오 장면에 대해 하나 이상의 비디오 스트림들을 생성하기 위해 사용될 수 있는 복수의 미디어 인코더들. 간략화를 위해, 비디오 인코더들은 도면에 표현되지 않는다.
- [0233] ○ 상이한 비트레이트들(즉, 상이한 표현들)의 동일한 오디오 및 비디오 스트림들의 상이한 인코딩들을 포함하는 복수의 오디오 및 비디오 적응 세트들을 저장하는 미디어 서버. 부가적으로, 미디어 서버는 모든 적응 세트들의 설명 정보를 포함하며, 그 설명 정보는 다음을 포함할 수 있다:
- [0235] ■ 생성된 모든 적응 세트들의 이용가능성;
- [0236] ■ 하나의 적응 세트와 하나의 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보; 이러한 방식으로, 각각의 적응 세트는 이용가능한 오디오 장면 중 하나와 연관될 수 있음;
- [0237] ■ (예를 들어, 완전한 오디오 장면 또는 단지 개별적인 오디오 오브젝트들을 포함할 수 있는) 각각의 오디오 장면 및/또는 뷰포인트의 "경계들"을 설명하는 정보. 하나의 오디오 장면의 경계들은, 예를 들어 구의 기하학적 좌표들(예를 들어, 중심 및 반경)로서 정의될 수 있음.
- [0238] ● 클라이언트 측에서, 시스템(클라이언트 시스템)은 다음 중 임의의 것을 포함할 수 있다:
- [0239] ○ 다음을 수신할 수 있는 수신단:
- [0240] ■ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보),
- [0241] ■ 모든 적응 세트들의 이용가능성에 관한 정보 및 하나의 적응 세트와 하나의 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보; 및/또는 (예를 들어, 완전한 오디오 장면 또는 단지 개별적인 오브젝트들을 포함할 수 있는) 각각의 오디오 장면 및/또는 뷰포인트의 "경계들"을 설명하는 정보. 예를 들어, 그러한 정보는 DASH

전달 환경의 경우, 미디어 프리젠테이션 설명(MPD) XML 선택스의 일부로서 제공될 수 있음.

- [0242] ○ (예를 들어, HMD에 기초한) 콘텐츠 소비를 위해 사용되는 미디어 소비 디바이스 측, 미디어 소비 디바이스는 또한, 사용자 위치 및/또는 배향 및/또는 움직임의 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보) 수집을 담당함.
- [0243] ○ 다음과 같이 구성될 수 있는 뷰포트 프로세서(1232):
- [0244] ■ 미디어 소비 디바이스 측으로부터, 사용자 위치 및/또는 배향 및/또는 움직임 방향을 포함할 수 있는 현재 뷰포인트에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보)를 수신함.
- [0245] ■ 메타데이터에 관한 정보 및 메타데이터에서 시그널링된 ROI(OMAF 구역에서와 같이 시그널링된 비디오 뷰포트들)를 수신함.
- [0246] ■ 수신단에서 이용가능한 모든 정보를 수신함.
- [0247] ■ 수신된 및/또는 이용가능한 메타데이터로부터 수신된 및/또는 도출된 모든 정보에 기초하여, 특정한 시간 순간에 어떤 오디오/비디오 뷰포인트가 재생되어야 하는지를 판단함. 예를 들어, 뷰포트 프로세서(1232)는 다음을 판단할 수 있다:
 - 하나의 완전한 오디오 장면이 재생될 것이라는 것,
 - 하나의 새로운 오디오 장면이 모든 이용가능한 오디오 장면들 중에서 생성되어야 한다는 것(예를 들어, 복수의 오디오 장면들로부터의 일부 오디오 엘리먼트들만이 재생될 것인 반면, 이들 오디오 장면들의 다른 나머지 오디오 엘리먼트들은 재생되지 않을 것),
 - 2개 이상의 오디오 장면들 사이의 전환이 재생되어야 한다는 것.
- [0251] ○ 뷰포트 프로세서(1232)로부터 수신된 정보에 기초하여, 수신단에 의해 수신된 정보에서 시그널링된 바와 같은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 이루어진 선택 부분(1230); 선택된 적응 세트들은 사용자의 현재 위치에서 재생되어야 하는 오디오 장면을 완전하게 설명함. 이러한 오디오 장면은 인코딩 측에서 정의된 바와 같이 하나의 완전한 오디오 장면일 수 있거나, 또는 새로운 오디오 장면은 모든 이용가능한 오디오 장면들 중에서 생성되어야 함.
- [0252] ■ 부가적으로, 2개 이상의 오디오 장면들 사이의 전환이 뷰포트 프로세서(1232)의 표시에 기초하여 발생하려고 하는 경우, 선택 부분(1230)은 수신단에 의해 수신된 정보에서 시그널링된 바와 같은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 구성될 수 있으며; 선택된 적응 세트들은 가까운 미래에 재생되도록 요구될 수 있는 오디오 장면을 완전하게 설명함(예를 들어, 사용자가 특정한 속도로 다음 오디오 장면의 방향으로 걷는다면, 다음 오디오 장면이 요구될 것이라고 예측될 수 있고, 다음 오디오 장면은 재생 전에 미리 선택됨).
- [0253] ■ 부가적으로, 이웃한 위치들에 대응하는 일부 적응 세트들은 먼저 더 낮은 비트레이트로 선택될 수 있고(즉, 더 낮은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택됨), 위치 변화들에 기초하여, 품질은 이들 특정 적응 세트들에 대해 더 높은 비트레이트를 선택함으로써 증가됨(즉, 더 높은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택됨).
- [0254] ○ 다음과 같이 구성될 수 있는 다운로드 및 스위칭 부분:
- [0255] ■ 선택 부분(1230)으로부터 수신된 표시에 기초하여, 미디어 서버(120)로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 요청함;
- [0256] ■ 미디어 서버(120)로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들(즉, 각각의 적응 세트 내에서 이용가능한 모든 표현들 중에서 하나의 표현)을 수신함;
- [0257] ■ 모든 수신된 오디오 스트림들로부터 메타데이터 정보를 추출함.
- [0258] ○ 다음과 같이 구성될 수 있는 메타데이터 프로세서(1236):
- [0259] ■ 수신된 오디오 스트림들에 관한 다운로드 및 스위칭 정보로부터, 수신된 각각의 오디오 스트림에 대응하는 오디오 메타데이터를 포함할 수 있는 정보,
- [0260] ■ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보를 포함할 수 있는 뷰포트 프로세서(1232)로부

터 수신된 정보에 기초하여, 각각의 오디오 스트림과 연관된 오디오 메타데이터를 프로세싱 및 조작하여:

- [0261] ● 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면을 조립하는 요구되는 오디오 엘리먼트들(152)을 선택/인에이블링시킴;
- [0262] ● 모든 오디오 스트림들의 단일 오디오 스트림으로의 병합을 허용함.
- [0263] ○ 모든 수신된 오디오 스트림들에 대응하는 수정된 및 프로세싱된 오디오 메타데이터를 포함할 수 있는 메타데이터 프로세서(1236)로부터 수신된 정보에 기초하여, 모든 선택된 오디오 스트림들을 하나의 오디오 스트림으로 병합하도록 구성될 수 있는 스트림 다중화기/병합기(1238),
- [0264] ○ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면의 재생을 위해 적어도 하나의 오디오 스트림을 수신 및 디코딩하도록 이루어진 미디어 디코더.
- [0265] 도 1c은 클라이언트 측에서, 예를 들어 다음을 부가적으로 또는 대안적으로 포함하는 도 1b에 설명된 시스템의 일부들을 구현할 수 있는 시스템(클라이언트 시스템)을 포함하는 시스템을 도시하며:
- [0266] ● (예를 들어, 복수의 오디오 엘리먼트들이 비활성화되어 있는) 메타데이터 프로세서(1236)에 의해 표시된 바와 같이 개별 오디오 스트림들을 디코딩하도록 구성될 수 있는 복수의 오디오 디코더들.
- [0267] ● 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 최종 오디오 장면을 재생하도록 구성될 수 있는 믹서/렌더러(1238)(즉 예를 들어, 그 특정 위치에서 가청적이지 않은 오디오 엘리먼트들 중 일부는 디스에이블링되거나 렌더링되지 않아야 함).
- [0268] **솔루션 2**
- [0269] 도 1d, 도 1e 및 도 1f은 (도 1a 및/또는 도 1b 및/또는 도 1c의 예들의 실시예들을 포함할 수 있는) 본 발명의 솔루션 2에 따른 예들을 참조한다: 유연한 적응 세트들을 갖는 별개의 위치들(뷰포인트들)에 대한 독립적인 적응 세트들.
- [0270] 사용자가 VR 환경에서 이동하는 경우, 오디오 장면(150)은 연속적인 방식으로 변화될 수 있다. 양호한 오디오 경험을 보장하기 위해, 특정한 시간 순간에 오디오 장면(150)을 조립하는 모든 오디오 엘리먼트들(152)은, 최종 오디오 장면을 생성하기 위해 위치 정보를 이용할 수 있는 미디어 디코더에 이용가능하게 되어야 할 수 있다.
- [0271] 복수의 미리-정의된 위치들에 대해 콘텐츠가 미리-인코딩되면, 시스템은, 오디오 장면들이 중첩되지 않고 사용자가 하나의 위치로부터 다음 위치로 "점프/스위칭"할 수 있다는 가정 하에서, 이들 특정 위치들에서 이들 오디오 장면들의 정확한 재생을 제공할 수 있다.
- [0272] 그러나, 사용자가 하나의 위치로부터 다음 위치로 "걷는" 경우들에서, 2개(또는 그 이상)의 오디오 장면들(150)로부터의 오디오 엘리먼트들(152)이 동시에 가청적일 수 있다. 이러한 사용 경우들에 대한 솔루션은 이전의 시스템 예들에서 제공되었으며, 여기서 (단일 미디어 디코더 또는 복수의 미디어 디코더를 갖는 다중화기 및 부가적인 믹서/렌더러(1238)를 사용하여) 복수의 오디오 스트림들을 디코딩하기 위해 제공되는 메커니즘들과 독립적으로, 완전한 오디오 장면들(150)을 설명하는 오디오 스트림들은 클라이언트/시스템(102)에 제공되어야 한다.
- [0273] 다음에서, 복수의 오디오 스트림들 사이에 공통 오디오 엘리먼트들(152)의 개념을 도입함으로써 최적화가 제공된다.
- [0274] 도 1d는, 상이한 장면들이 적어도 하나의 오디오 엘리먼트(오디오 오브젝트, 사운드 소스 등)를 공유하는 일 예를 도시한다. 따라서, 클라이언트(102)는, 예를 들어 하나의 장면 A에만 연관되고(예를 들어, 사용자가 현재 있는 환경에 연관되고) 오브젝트들(152A)에 연관된 하나의 주요 스트림(106A), 및 상이한 장면 B에 의해 공유되고 오브젝트들(152B)에 연관된 하나의 보조 스트림(106B)(예를 들어, 오브젝트들(152B)을 공유하는 이웃한 또는 인접한 스트림 B와 사용자가 현재 있는 장면 A 사이의 경계 내의 스트림)을 수신할 수 있다.
- [0275] 따라서, 도 1d에 도시된 바와 같이:
- [0276] ● 수 개의 독립적이거나 중첩하는 오디오 장면들이 복수의 오디오 스트림들로 인코딩된다. 오디오 스트림들(106)은 다음과 같은 방식으로 생성된다:
- [0277] ○ 각각의 오디오 장면(150)에 대해, 하나의 주요 스트림은, 개개의 오디오 장면의 일부이지만 임의의 다른 오

디오 장면의 일부가 아닌 오디오 엘리먼트들(152)만을 포함함으로써 생성될 수 있고; 그리고/또는

- [0278] ○ 오디오 엘리먼트들(152)을 공유하는 모든 오디오 장면들(150)에 대해, 공통 오디오 엘리먼트들(152)은 오디오 장면들 중 오직 하나와만 연관된 보조 오디오 스트림들로만 인코딩될 수 있고, 다른 오디오 장면들과의 연관을 표시하는 적절한 메타데이터 정보가 생성되고, 또는 다르게 언급하면, 부가적인 메타데이터는, 일부 오디오 스트림들이 복수의 오디오 장면들과 함께 사용될 수 있는 가능성을 표시하고; 그리고/또는
- [0279] ○ 사용 경우에 의존하여, 부가적인 보조 스트림들이 생성될 수 있다(예를 들어, 상이한 언어들을 포함하는 일부 오디오 오브젝트들은 효율적인 전달을 위해 독립적인 스트림들로 인코딩될 수 있음).
- [0280] ○ 제공된 실시예에서:
- [0281] ■ 오디오 장면 A는 다음으로 인코딩된다:
- [0282] ● 메인 오디오 스트림(A1, 106A),
- [0283] ● 보조 오디오 스트림(A2, 106B),
- [0284] ● 오디오 장면 A로부터의 일부 오디오 엘리먼트들(152B)이 이들 오디오 스트림 A가 아니라 상이한 오디오 장면(오디오 장면 B)에 속하는 보조 스트림 A2(106B)로 인코딩된다는 것을 표시할 수 있는 메타데이터 정보.
- [0285] ■ 오디오 장면 B는 다음으로 인코딩된다:
- [0286] ● 메인 오디오 스트림(B1, 106C),
- [0287] ● 보조 오디오 스트림(B2),
- [0288] ● 보조 오디오 스트림(B3),
- [0289] ● 오디오 스트림 B2로부터의 오디오 엘리먼트들(152B)이 오디오 장면 A에 또한 속하는 공통 오디오 엘리먼트들(152B)이라는 것을 표시할 수 있는 메타데이터 정보.
- [0290] ■ 오디오 장면 C는 3개의 스트림들(C1, C2 및 C3)로 인코딩된다.
- [0291] ● 오디오 스트림들(106)(106A, 106B, 106C 등)은 상이한 비트레이트들(즉, 상이한 표현들)로 인코딩되어, 예를 들어 네트워크 연결에 의존하여 효율적인 비트레이트 적응을 허용할 수 있다(즉, 고속 연결을 사용하는 사용자들에 대해서는 높은 비트레이트 코딩된 버전이 전달되는 반면, 더 낮은 속도의 네트워크 연결을 갖는 사용자들에 대해서는 더 낮은 비트레이트 버전이 전달됨).
- [0292] ● 오디오 스트림들(106)은 미디어 서버(120) 상에 저장되며, 여기서 각각의 오디오 스트림에 대해, 상이한 비트레이트들(즉, 상이한 표현들)의 상이한 인코딩들은 하나의 적응 세트로 그룹화되는 데, 적절한 데이터가 모든 생성된 적응 세트들의 이용가능성을 시그널링한다. (동일한 오디오 신호들에 연관되지만 상이한 비트레이트들 및/또는 품질들 및/또는 해상도들에 있는 스트림들의 복수의 표현들이 동일한 적응 세트에 존재할 수 있음).
- [0293] ● 적응 세트들에 부가적으로, 미디어 서버(120)는 각각의 오디오 장면의 위치 "경계들" 및 (예를 들어 완전한 오디오 장면 또는 단지 개별적인 오브젝트들을 포함할 수 있는) 각각의 적응 세트에 대한 그의 관계에 관한 정보를 수신할 수 있다. 이러한 방식으로, 각각의 적응 세트는 이용가능한 오디오 장면들(150) 중 하나 이상과 연관될 수 있다. 하나의 오디오 장면의 경계들은, 예를 들어 구의 기하학적 좌표들(예를 들어, 중심 및 반경)로서 정의될 수 있다.
- [0294] ○ 각각의 적응 세트는, 사운드 장면 또는 오디오 엘리먼트들(152)이 활성화된 위치들에 관한 설명 정보를 또한 포함할 수 있다. 예를 들어, 하나의 보조 스트림(예를 들어, A2, 106B)이 하나 또는 수 개의 오브젝트들을 포함하면, 적응 세트는, 오브젝트들이 가칭적인 위치들(예를 들어, 구의 중심의 좌표들 및 반경)과 같은 정보를 포함할 수 있다.
- [0295] ○ 부가적으로 또는 대안적으로, 각각의 적응 세트(예를 들어, 장면 B에 연관된 적응 세트)는 하나의 오디오 장면(예를 들어, B)으로부터의 오디오 엘리먼트들(예를 들어, 152B)이 상이한 오디오 장면(예를 들어, A)에 속하는 오디오 스트림들(예를 들어, 106B)로 (또한 또는 부가적으로) 인코딩된다는 것을 표시할 수 있는 설명 정보(예를 들어, 메타데이터)를 포함할 수 있다.
- [0296] ● 미디어 서버(120)는 각각의 적응 세트와 연관된 위치 "경계들"에 관한 정보를 시스템(102)(클라이언트), 예를 들어 DASH 클라이언트에 제공할 수 있다. 예를 들어, 이것은 DASH 전달 환경의 경우, 미디어 프리젠테이션

설명(MPD) XML 신택스로 임베딩될 수 있다.

- [0297] ● 시스템(102)(클라이언트)은 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보)를 수신할 수 있다.
- [0298] ● 시스템(102)(클라이언트)은 각각의 적응 세트에 관한 정보를 수신할 수 있고, 그 정보 및/또는 사용자 위치 및/또는 배향 및/또는 움직임 방향(또는, 예를 들어 x,y,z 좌표들 및/또는 요, 피치, 롤 값들을 포함하는, 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보)에 기초하여, 시스템(102)(클라이언트)은 사용자(140)의 현재 위치에서 재생되어야 하는 오디오 장면(150)을 완전하게 또는 부분적으로 설명하는 하나 이상의 적응 세트들을 선택할 수 있다.
- [0299] ● 시스템(102)(클라이언트)은 하나 이상의 적응 세트들을 요청할 수 있다:
- [0300] ○ 게다가, 시스템(102)(클라이언트)은 하나 초과 오디오 장면(150)을 완전하게 또는 부분적으로 설명하는 하나 이상의 적응 세트들을 선택하고, 하나 초과 오디오 장면(150)에 대응하는 오디오 스트림들(106)을 사용하여, 사용자(140)의 현재 위치에서 재생될 새로운 오디오 장면(150)을 생성할 수 있다.
- [0301] ○ 오디오 엘리먼트들(152)이 복수의 오디오 장면들(150)의 일부라는 것을 표시하는 메타데이터에 기초하여, 공통 오디오 엘리먼트들(152)은 각각의 완전한 오디오 장면에 대해 한번씩, 그들을 2회 요청하는 대신, 새로운 오디오 장면을 생성하기 위해 단지 한번 요청될 수 있다.
- [0302] ○ 일단 오디오 스트림들이 클라이언트 시스템(102)에 이용가능하면, 예들에서, 하나 또는 복수의 미디어 디코더(들)(104)는 개별 오디오 스트림들을 디코딩하는 데 사용될 수 있고, 그리고/또는 부가적인 믹서/렌더러는 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 최종 오디오 장면을 재생하는 데 사용될 수 있다(즉 예를 들어, 그 특정 위치에서 가청적이지 않은 오디오 엘리먼트들 중 일부는 디스에이블링되거나 렌더링되지 않아야 함).
- [0303] ○ 대안적으로 또는 부가적으로, 메타데이터 프로세서는 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 모든 오디오 스트림들과 연관된 오디오 메타데이터를 조작하는 데 사용되어:
- [0304] ■ 새로운 오디오 장면을 조립하는 요구되는 오디오 엘리먼트들(152)(152A 내지 152C)을 선택/인에이블링하고; 그리고/또는
- [0305] ■ 그리고 모든 오디오 스트림들의 단일 오디오 스트림으로의 병합을 허용할 수 있다.
- [0306] ● 미디어 서버(120)는 요구되는 적응 세트들을 전달할 수 있다.
- [0307] ● 대안적으로, 시스템(102)(클라이언트)은 사용자(140) 위치설정에 관한 정보를 미디어 서버(120)에 제공하고, 미디어 서버는 요구되는 적응 세트들에 관한 표시를 제공한다.
- [0308] 도 1e는 다음을 포함하는 그러한 시스템의 다른 예시적인 구현을 도시한다:
- [0309] ● 인코딩 측에서,
- [0310] ○ 하나의 뷰포인트의 하나의 사운드 장면 부분과 연관된 하나 이상의 이용가능한 오디오 장면(150)으로부터 오디오 엘리먼트들(152)을 임베딩한 하나 이상의 오디오 스트림들(106)을 생성하기 위해 사용될 수 있는 복수의 미디어 인코더들(154),
- [0311] ■ 각각의 오디오 장면(150)에 대해, 하나의 주요 스트림은, 개개의 오디오 장면(150)의 일부이지만 임의의 다른 오디오 장면의 일부가 아닌 오디오 엘리먼트들(152)만을 포함함으로써 생성될 수 있다.
- [0312] ■ 부가적인 보조 스트림들이 동일한 오디오 장면에 대해 생성될 수 있다(예를 들어, 상이한 언어들을 포함하는 일부 오디오 오브젝트들은 효율적인 전달을 위해 독립적인 스트림들로 인코딩될 수 있음).
- [0313] ■ 다음을 포함하는 부가적인 보조 스트림들이 생성될 수 있다:
- [0314] ● 하나 초과 오디오 장면(150)에 공통적인 오디오 엘리먼트들(152),
- [0315] ● 공통 오디오 엘리먼트들(152)을 공유하는 다른 모든 오디오 장면들(150)과 이러한 보조 스트림의 연관을 표시하는 메타데이터. 또는 다르게 언급하면, 메타데이터는, 일부 오디오 스트림들이 복수의 오디오 장면들과 함께 사용될 수 있는 가능성을 표시함.

- [0316] ○ 하나의 뷰포인트의 하나의 비디오 장면 부분과 연관된 각각의 이용가능한 비디오 장면에 대해 하나 이상의 비디오 스트림들을 생성하기 위해 사용될 수 있는 복수의 미디어 인코더들. 간략화를 위해, 비디오 인코더들은 도면에 표현되지 않는다.
- [0317] ○ 상이한 비트레이트들(즉, 상이한 표현들)의 동일한 오디오 및 비디오 스트림들의 상이한 인코딩들을 포함하는 복수의 오디오 및 비디오 적응 세트들을 저장하는 미디어 서버(120). 부가적으로, 미디어 서버(120)는 모든 적응 세트들의 설명 정보를 포함하며, 그 설명 정보는 다음을 포함할 수 있다:
- [0318] ■ 생성된 모든 적응 세트들의 이용가능성;
- [0319] ■ 하나의 적응 세트와 하나의 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보; 이러한 방식으로, 각각의 적응 세트는 이용가능한 오디오 장면 중 하나와 연관될 수 있음;
- [0320] ■ (예를 들어, 완전한 오디오 장면 또는 단지 개별적인 오디오 오브젝트들을 포함할 수 있는) 각각의 오디오 장면 및/또는 뷰포인트의 "경계들"을 설명하는 정보. 하나의 오디오 장면의 경계들은, 예를 들어 구의 기하학적 좌표들(예를 들어, 중심 및 반경)로서 정의될 수 있음.
- [0321] ■ 적어도 하나의 공통 오디오 엘리먼트를 공유하는 하나 초과와 오디오 장면과 하나의 적응 세트의 연관을 표시하는 정보.
- [0322] ● 클라이언트 측에서, 시스템(클라이언트 시스템)은 다음 중 임의의 것을 포함할 수 있다:
- [0323] ○ 다음을 수신할 수 있는 수신단:
- [0324] ■ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보),
- [0325] ■ 모든 적응 세트들의 이용가능성에 관한 정보 및 하나의 적응 세트와 하나의 오디오 장면 및/또는 뷰포인트의 연관을 설명하는 정보; 및/또는 (예를 들어, 완전한 오디오 장면 또는 단지 개별적인 오브젝트들을 포함할 수 있는) 각각의 오디오 장면 및/또는 뷰포인트의 "경계들"을 설명하는 정보. 예를 들어, 그러한 정보는 DASH 전달 환경의 경우, 미디어 프리젠테이션 설명(MPD) XML 섹스의 일부로서 제공될 수 있음.
- [0326] ■ 적어도 하나의 공통 오디오 엘리먼트를 공유하는 하나 초과와 오디오 장면과 하나의 적응 세트의 연관을 표시하는 정보.
- [0327] ○ (예를 들어, HMD에 기초한) 콘텐츠 소비를 위해 사용되는 미디어 소비 디바이스 측. 미디어 소비 디바이스는 또한, 사용자 위치 및/또는 배향 및/또는 움직임의 방향에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보) 수집을 담당함.
- [0328] ○ 다음과 같이 구성될 수 있는 뷰포트 프로세서(1232):
- [0329] ■ 미디어 소비 디바이스 측으로부터, 사용자 위치 및/또는 배향 및/또는 움직임 방향을 포함할 수 있는 현재 뷰포인트에 관한 정보(또는 사용자의 액션들에 의해 트리거링된 변화들을 특성화하는 임의의 정보)를 수신함.
- [0330] ■ 메타데이터에 관한 정보 및 메타데이터에서 시그널링된 ROI(OMAF 구역에서와 같이 시그널링된 비디오 뷰포트들)를 수신함.
- [0331] ■ 수신단에서 이용가능한 모든 정보를 수신함.
- [0332] ■ 수신된 및/또는 이용가능한 메타데이터로부터 수신된 및/또는 도출된 모든 정보에 기초하여, 특정한 시간 순간에 어떤 오디오/비디오 뷰포인트가 재생되어야 하는지를 판단함. 예를 들어, 뷰포트 프로세서(1232)는 다음을 판단할 수 있다:
- [0333] ● 하나의 완전한 오디오 장면이 재생될 것이라는 것,
- [0334] ● 하나의 새로운 오디오 장면이 모든 이용가능한 오디오 장면들 중에서 생성되어야 한다는 것(예를 들어, 복수의 오디오 장면들로부터의 일부 오디오 엘리먼트들만이 재생될 것인 반면, 이들 오디오 장면들의 다른 나머지 오디오 엘리먼트들은 재생되지 않을 것),
- [0335] ● 2개 이상의 오디오 장면들 사이의 전환이 재생되어야 한다는 것.
- [0336] ○ 뷰포트 프로세서(1232)로부터 수신된 정보에 기초하여, 수신단에 의해 수신된 정보에서 시그널링된 바와 같

은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 이루어진 선택 부분(1230); 선택된 적응 세트들은 사용자의 현재 위치에서 재생되어야 하는 오디오 장면을 완전하게 또는 부분적으로 설명함. 이러한 오디오 장면은 인코딩 측에서 정의된 바와 같이 하나의 완전한 또는 부분적으로 완전한 오디오 장면일 수 있거나, 또는 새로운 오디오 장면은 모든 이용가능한 오디오 장면들 중에서 생성되어야 함.

[0337] ■ 부가적으로, 오디오 엘리먼트들(152)이 하나 초과와 오디오 장면에 속하는 경우, 적어도 하나의 적응 세트들은, 동일한 오디오 엘리먼트들(152)을 포함하는, 하나 초과와 오디오 장면과 적어도 하나의 적응 세트의 연관을 표시하는 정보에 기초하여 선택됨.

[0338] ■ 부가적으로, 2개 이상의 오디오 장면들 사이의 전환이 뷰포트 프로세서(1232)의 표시에 기초하여 발생하려고 하는 경우, 선택 부분(1230)은 수신단에 의해 수신된 정보에서 시그널링된 바와 같은 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 선택하도록 구성될 수 있으며; 선택된 적응 세트들은 가까운 미래에 재생되도록 요구될 수 있는 오디오 장면을 완전하게 설명함(예를 들어, 사용자가 특정한 속도로 다음 오디오 장면의 방향으로 걷는다면, 다음 오디오 장면이 요구될 것이라고 예측될 수 있고, 다음 오디오 장면은 재생 전에 미리 선택됨).

[0339] ■ 부가적으로, 이웃한 위치들에 대응하는 일부 적응 세트들은 먼저 더 낮은 비트레이트로 선택될 수 있고(즉, 더 낮은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택됨), 위치 변화들에 기초하여, 품질은 이들 특정 적응 세트들에 대해 더 높은 비트레이트를 선택함으로써 증가됨(즉, 더 높은 비트레이트로 인코딩된 표현은 하나의 적응 세트 내의 이용가능한 표현들 중에서 선택됨).

[0340] ○ 다음과 같이 구성될 수 있는 다운로드 및 스위칭 부분:

[0341] ■ 선택 부분(1230)으로부터 수신된 표시에 기초하여, 미디어 서버(120)로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들을 요청함;

[0342] ■ 미디어 서버(120)로부터의 이용가능한 적응 세트들 중에서 하나 이상의 적응 세트들(즉, 각각의 적응 세트 내에서 이용가능한 모든 표현들 중에서 하나의 표현)을 수신함;

[0343] ■ 모든 수신된 오디오 스트림들로부터 메타데이터 정보를 추출함.

[0344] ○ 다음과 같이 구성될 수 있는 메타데이터 프로세서(1236):

[0345] ■ 수신된 오디오 스트림들에 관한 다운로드 및 스위칭 정보로부터, 수신된 각각의 오디오 스트림에 대응하는 오디오 메타데이터를 포함할 수 있는 정보,

[0346] ■ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보를 포함할 수 있는 뷰포트 프로세서(1232)로부터 수신된 정보에 기초하여, 각각의 오디오 스트림과 연관된 오디오 메타데이터를 프로세싱 및 조작하여:

[0347] ● 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면을 조립하는 요구되는 오디오 엘리먼트들(152)을 선택/인에이블링시킴;

[0348] ● 모든 오디오 스트림들의 단일 오디오 스트림으로의 병합을 허용함.

[0349] ○ 모든 수신된 오디오 스트림들에 대응하는 수정된 및 프로세싱된 오디오 메타데이터를 포함할 수 있는 메타데이터 프로세서(1236)로부터 수신된 정보에 기초하여, 모든 선택된 오디오 스트림들을 하나의 오디오 스트림으로 병합하도록 구성될 수 있는 스트림 다중화기/병합기(1238),

[0350] ○ 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 뷰포트 프로세서(1232)에 의해 표시된 바와 같이 새로운 오디오 장면의 재생을 위해 적어도 하나의 오디오 스트림을 수신 및 디코딩하도록 이루어진 미디어 디코더.

[0351] 도 1f는 클라이언트 측에서, 예를 들어 다음을 부가적으로 또는 대안적으로 포함하는 도 5에 설명된 시스템의 일부들을 구현할 수 있는 시스템(클라이언트 시스템)을 포함하는 시스템을 도시하며:

[0352] ● (예를 들어, 복수의 오디오 엘리먼트들이 비활성화되어 있는) 메타데이터 프로세서(1236)에 의해 표시된 바와 같이 개별 오디오 스트림들을 디코딩하도록 구성될 수 있는 복수의 오디오 디코더들.

[0353] ● 사용자 위치 및/또는 배향 및/또는 움직임 방향에 관한 정보에 기초하여 최종 오디오 장면을 재생하도록 구성될 수 있는 믹서/렌더러(1238)(즉 예를 들어, 그 특정 위치에서 가청적이지 않은 오디오 엘리먼트들 중 일부

는 디스에이블링되거나 렌더링되지 않아야 함).

[0354] 파일 재생을 위한 파일 포맷 업데이트들

[0355] 파일 포맷 사용 경우에 있어서, 복수의 메인 및 보조 스트림들은 별개의 트랙(Track)들로서 단일 ISOBMFF 파일로 캡슐화될 수 있다. 그러한 파일의 단일 트랙은 이전에 언급된 바와 같이 단일 오디오 엘리먼트를 표현할 것이다. 정확한 재생을 위해 필요한 정보를 포함하는 이용가능한 어떠한 MPD도 존재하지 않으므로, 예를 들어 트랙 및 영화 레벨에 대해 특정 파일 포맷 박스 또는 특정 파일 포맷 박스들을 제공/도입함으로써, 정보가 파일 포맷 레벨에 대해 제공될 필요가 있다. 사용-경우에 의존하여, 캡슐화된 오디오 장면들의 정확한 렌더링을 허용하는 데 필요한 상이한 정보가 존재하지만, 정보의 다음의 세트가 기본적이고, 그에 따라 항상 존재해야 한다:

[0356] ● 포함된 오디오 장면에 관한 정보, 예를 들어 "위치 경계들",

[0357] ● 모든 이용가능한 오디오 엘리먼트들, 특히 어느 오디오 엘리먼트가 어느 트랙으로 캡슐화되는지에 관한 정보,

[0358] ● 캡슐화된 오디오 엘리먼트들의 위치에 관한 정보,

[0359] ● 하나의 오디오 장면에 속하는 모든 오디오 엘리먼트들의 리스트, 오디오 엘리먼트는 복수의 오디오 장면에 속할 수 있음.

[0360] 이러한 정보를 이용하면, 부가적인 메타데이터 프로세서 및 공유된 인코딩을 갖는 사용 경우를 포함하는 모든 언급된 사용 경우들이 또한 파일 기반 환경에서 작동되어야 한다.

[0361] 위의 예들에 대한 추가적인 고려사항들

[0362] 예들(예를 들어, 도 1a 내지 도 1f 중 적어도 하나)에서, 적어도 하나의 장면은 적어도 하나의 오디오 엘리먼트(오디오 소스(152))에 연관될 수 있으며, 각각의 오디오 엘리먼트는, 오디오 엘리먼트가 가청적인 시각적 환경 내의 위치 및/또는 영역에 연관되어, 상이한 오디오 스트림들이 장면 내의 상이한 사용자의 위치들 및/또는 뷰포트들 및/또는 머리 배향들 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 대해 서버 시스템(120)으로부터 클라이언트 시스템(102)으로 제공된다.

[0363] 예들에서, 클라이언트 시스템(102)은, 오디오 스트림(예를 들어, A1, A2)의 적어도 하나의 오디오 엘리먼트(152) 및/또는 하나의 적응 세트가 장면 내의 현재 사용자의 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치의 존재 시에 재생될지 여부를 판단하도록 구성될 수 있으며, 시스템(102)은 현재 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 이루어진다.

[0364] 예들에서, 클라이언트 시스템(예를 들어, 102)은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)에 적어도 기초하여, 오디오 스트림의 적어도 하나의 오디오 엘리먼트(152) 및/또는 하나의 적응 세트가 관련있게 될지 그리고/또는 가청적있게 될지를 예측적으로 판단하도록 구성될 수 있으며, 시스템은, 장면 내의 예측된 사용자의 움직임 및/또는 상호작용 이전에 특정한 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림 및/또는 적응 세트를 요청 및/또는 수신하도록 이루어지고, 시스템은, 수신될 경우, 장면 내의 사용자의 움직임 및/또는 상호작용 이후 특정한 사용자의 가상 위치에서 적어도 하나의 오디오 엘리먼트 및/또는 오디오 스트림을 재생하도록 이루어진다. 예를 들어, 위의 도 8a 및 도 8b를 참조한다. 일부 예들에서, 시스템(102 또는 120)의 동작들 중 적어도 하나는 예측 및/또는 통계 및/또는 집계 데이터에 기초하여 수행될 수 있다.

[0365] 예들에서, 클라이언트 시스템(예를 들어, 102)은, 장면 내의 사용자의 움직임 및/또는 상호작용 이전에 사용자의 가상 위치에서 더 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 오디오 엘리먼트(예를 들어, 152)를 요청 및/또는 수신하도록 구성될 수 있으며, 여기서 시스템은, 장면 내의 사용자의 움직임 및/또는 상호작용 이후 사용자의 가상 위치에서 더 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 오디오 엘리먼트를 요청 및/또는 수신하도록 이루어진다. 예를 들어, 도 7b를 참조한다.

[0366] 예들에서, 적어도 하나의 오디오 엘리먼트는 적어도 하나의 장면에 연관될 수 있고, 적어도 하나의 오디오 엘리먼트는 장면에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관되며, 시스템은, 장면 내의 각각의 사용자의 가상 위치에서의 오디오 엘리먼트들의 관련성 및/또는 가청성 레벨에 기초하여 오디오 엘리먼트들에 대해 상이한 비트레이트들 및/또는 품질 레벨들로 상이한 스트림들을 요청하도록 이루어지고, 시스템은, 현재 사용자의

가상 위치에서 더 관련있고 그리고/또는 더 가청적인 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 오디오 스트림을 요청하고 그리고/또는 현재 사용자의 가상 위치에서 덜 관련있고 덜 가청적인 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 오디오 스트림을 요청하도록 이루어진다. 일반적인 측면들에서, 도 7a를 참조한다. 또한, 도 2a 및 도 2b(여기서, 더 많은 관련있는 및/또는 가청 소스들은 사용자에게 더 가까울 수 있음), 도 3(여기서, 사용자가 위치(x_1)에 있는 경우, 더 관련있는 및/또는 가청 소스는 장면(150a)의 소스이고, 사용자가 위치(x_3)에 있는 경우, 더 관련있는 및/또는 가청 소스는 장면(150b)의 소스임), 도 4(여기서, 시간 순간(t_2)에서, 더 관련있는 및/또는 가청 소스들은 제1 장면의 소스들일 수 있음), 도 6(여기서, 더 가청적인 소스들은 사용자에게 의해 전방에서 보여지는 소스들일 수 있음)을 참조한다.

[0367] 예들에서, 적어도 하나의 오디오 엘리먼트(152)는 장면(152)에 연관되고, 각각의 오디오 엘리먼트는 장면(152)에 연관된 시각적 환경 내의 위치 및/또는 영역에 연관되며, 클라이언트 시스템(102)은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터(110)를 서버 시스템(120)에 주기적으로 전송하여, 적어도 하나의 오디오 엘리먼트(152)에 더 가까운 위치의 경우, 더 높은 비트레이트 및/또는 품질의 스트림이 서버로부터 제공되고, 적어도 하나의 오디오 엘리먼트(152)로부터 더 멀리있는 위치의 경우, 더 낮은 비트레이트 및/또는 품질의 스트림이 서버로부터 제공되도록 이루어진다. 예를 들어, 도 2a 및 도 2b를 참조한다.

[0368] 예들에서, 복수의 장면들(예를 들어, 150A, 150B)이 인접한 및/또는 이웃한 환경들과 같은 복수의 시각적 환경들에 대해 정의될 수 있어서, 제1의 현재 장면(예를 들어, 150A)에 연관된 제1 스트림들이 제공되고, 제2의 추가적인 장면(예를 들어, 150B)으로의 사용자의 전환(150AB)의 경우에는, 제1 장면(150A)에 연관된 스트림들 및 제2 장면(150B)에 연관된 제2 스트림들 둘 모두가 제공된다. 예를 들어, 도 3을 참조한다.

[0369] 예들에서, 복수의 장면들이 제1 및 제2 시각적 환경들에 대해 정의되며, 제1 및 제2 환경들은 인접한 및/또는 이웃한 환경들이고, 제1 장면(150A)에 연관된 제1 스트림들은, 사용자의 가상 위치가 제1 장면(150A)에 연관된 제1 환경에 있는 경우 제1 장면(150A)의 재생을 위해 서버로부터 제공되고, 제2 장면(150B)에 연관된 제2 스트림들은, 사용자의 가상 위치가 제2 장면(150B)에 연관된 제2 환경에 있는 경우 제2 장면(150B)의 재생을 위해 서버로부터 제공되며, 제1 장면(150A)에 연관된 제1 스트림들 및 제2 장면(150B)에 연관된 제2 스트림들 둘 모두는 사용자의 가상 위치가 제1 장면(150A)과 제2 장면(150B) 사이의 전환 위치에 있는 경우 제공된다. 예를 들어, 도 3을 참조한다.

[0370] 예들에서, 사용자가 제1 장면(150A)에 연관된 제1 환경에 있는 경우 제1 장면(150A)에 연관된 제1 스트림들은 더 높은 비트레이트 및/또는 품질로 획득되는 반면, 사용자가 제1 장면(150A)으로부터 제2 장면(150B)으로의 전환 위치에 있는 경우 제2 환경에 연관된 제2 장면 환경에 연관된 제2 스트림들은 더 낮은 비트레이트 및/또는 품질로 획득되며, 사용자가 제1 장면(150A)으로부터 제2 장면(150B)으로의 전환 위치에 있는 경우 제1 장면(150A)에 연관된 제1 스트림들은 더 낮은 비트레이트 및/또는 품질로 획득되고, 제2 장면(150B)에 연관된 제2 스트림들은 더 높은 비트레이트 및/또는 품질로 획득된다. 이것은, 예를 들어 도 3의 경우일 수 있다.

[0371] 예들에서, 복수의 장면들(예를 들어, 150A, 150B)은 복수의 시각적 환경들(예를 들어, 인접한 환경들)에 대해 정의되며, 시스템(102)은 현재 장면(150A)에 연관된 스트림들을 더 높은 비트레이트 및/또는 품질로 요청 및/또는 획득하고, 제2 장면(150B)에 연관된 스트림들을 더 낮은 비트레이트 및/또는 품질로 요청 및/또는 획득할 수 있다. 예를 들어, 도 4를 참조한다.

[0372] 예들에서, 복수의 N개의 오디오 엘리먼트들이 정의되며, 이들 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 큰 경우, N개의 오디오 엘리먼트들은 N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또는 영역에 연관된 더 작은 수 M개의 오디오 엘리먼트들($M < N$)을 획득하도록 프로세싱되어, N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 작은 경우 N개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림을 시스템에 제공하거나, 또는 N개의 오디오 엘리먼트들의 위치 또는 영역에 대한 사용자의 거리가 미리 결정된 임계치보다 큰 경우 M개의 오디오 엘리먼트들에 연관된 적어도 하나의 오디오 스트림을 시스템에 제공한다. 예를 들어, 도 1g를 참조한다.

[0373] 예들에서, 적어도 하나의 시각적 환경 장면은 적어도 하나의 복수의 N개의 오디오 엘리먼트들($N \geq 2$)에 연관되며, 각각의 오디오 엘리먼트는 시각적 환경 내의 위치 및/또는 영역에 연관되고, 적어도, 적어도 하나의 복수의 N개의 오디오 엘리먼트들은 높은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공될 수 있고, 적어도, 적어도 하나의 복수의 N개의 오디오 엘리먼트들은 낮은 비트레이트 및/또는 품질 레벨로 적어도 하나의 표현에서 제공되고, 적어도 하나의 표현은, N개의 오디오 엘리먼트들의 위치 또는 영역에 가까운 위치 또

는 영역에 연관된 더 작은 수 M 개의 오디오 엘리먼트들($M < N$)을 획득하도록 N 개의 오디오 엘리먼트들을 프로세싱함으로써 획득되고, 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 더 관련있고 그리고/또는 더 가청적인 경우 오디오 엘리먼트들에 대해 더 높은 비트레이트 및/또는 품질 레벨로 표현을 요청하도록 이루어지고, 시스템은, 오디오 엘리먼트들이 장면 내의 현재 사용자의 가상 위치에서 덜 관련있고 그리고/또는 덜 가청적인 경우 오디오 엘리먼트들에 대해 더 낮은 비트레이트 및/또는 품질 레벨로 표현을 요청하도록 이루어진다. 예를 들어, 도 1g을 참조한다.

[0374] 예들에서, 사용자의 거리 및/또는 관련성 및/또는 가청성 레벨 및/또는 각도 배향이 미리 결정된 임계치보다 낮은 경우, 상이한 스트림들이 상이한 오디오 엘리먼트들에 대해 획득된다. 예를 들어, 도 1g을 참조한다.

[0375] 예들에서, 상이한 오디오 엘리먼트들이 상이한 뷰포트들에서 제공되어, 하나의 제1 오디오 엘리먼트가 현재 뷰포트 내에 있는 경우, 제1 오디오 엘리먼트는 뷰포트 내에 있지 않은 제2 오디오 엘리먼트보다 더 높은 비트레이트로 획득된다. 예를 들어, 도 6을 참조한다.

[0376] 예들에서, 적어도 2개의 시각적 환경 장면들이 정의되며, 적어도 하나의 제1 및 제2 오디오 엘리먼트들은 제1 시각적 환경에 연관된 제1 장면에 연관되고, 적어도 하나의 제3 오디오 엘리먼트는 제2 시각적 환경에 연관된 제2 장면에 연관되고, 시스템(102)은 적어도 하나의 제2 오디오 엘리먼트가 제2 시각적 환경 장면과 부가적으로 연관된다는 것을 설명하는 메타데이터를 획득하도록 이루어지고, 시스템은, 사용자의 가상 위치가 제1 시각적 환경에 있는 경우 적어도 하나의 제1 및 제2 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어지고, 시스템은, 사용자의 가상 위치가 제2 시각적 환경 장면에 있는 경우 적어도 하나의 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어지며, 시스템은, 사용자의 가상 위치가 제1 시각적 환경 장면과 제2 시각적 환경 장면 사이의 전환에 있는 경우 적어도 하나의 제1 및 제2 및 제3 오디오 엘리먼트들을 요청 및/또는 수신하도록 이루어진다. 예를 들어, 도 1d를 참조한다. 이것은 또한 도 3에 적용될 수 있다.

[0377] 예들에서, 적어도 하나의 제1 오디오 엘리먼트가 적어도 하나의 오디오 스트림 및/또는 적응 세트에서 제공될 수 있고, 적어도 하나의 제2 오디오 엘리먼트가 적어도 하나의 제2 오디오 스트림 및/또는 적응 세트에서 제공되고, 적어도 하나의 제3 오디오 엘리먼트가 적어도 하나의 제3 오디오 스트림 및/또는 적응 세트에서 제공되며, 적어도 제1 시각적 환경 장면은 적어도 하나의 제1 및 제2 오디오 스트림들 및/또는 적응 세트들을 요구하는 완전한 장면으로서 메타데이터에 의해 설명되고, 제2 시각적 환경 장면은 적어도 제1 시각적 환경 장면과 연관된 적어도 하나의 제3 오디오 스트림 및/또는 적응 세트 및 적어도 하나의 제2 오디오 스트림 및/또는 적응 세트들을 요구하는 불완전한 장면으로서 메타데이터에 의해 설명되며, 시스템은, 사용자의 가상 위치가 제2 시각적 환경에 있는 경우, 제1 시각적 환경에 속하는 제2 오디오 스트림 및 제2 시각적 환경과 연관된 제3 오디오 스트림을 새로운 단일 스트림으로 병합하는 것을 허용하기 위해 메타데이터를 조작하도록 이루어진 메타데이터 프로세서를 포함한다. 예를 들어, 도 1b, 도 1c, 도 1e 및 도 1f을 참조한다.

[0378] 예들에서, 시스템(102)은, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 메타데이터를 조작하도록 이루어진 메타데이터 프로세서(예를 들어, 1236)를 포함할 수 있다.

[0379] 예들에서, 메타데이터 프로세서(예를 들어, 1236)는, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에 기초하여, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 인에이블링 및/또는 디스에이블링시키도록 구성될 수 있으며, 메타데이터 프로세서는, 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 오디오 엘리먼트가 더 이상 재생되지 않을 것이라고 시스템이 판단하는 경우, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 디스에이블링시키도록 구성될 수 있고, 메타데이터 프로세서는, 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터의 결과로서 오디오 엘리먼트가 재생될 것이라고 시스템이 판단하는 경우, 적어도 하나의 오디오 디코더 이전에 적어도 하나의 오디오 스트림에서 적어도 하나의 오디오 엘리먼트를 인에이블링시키도록 구성될 수 있다.

[0380] 서버 측

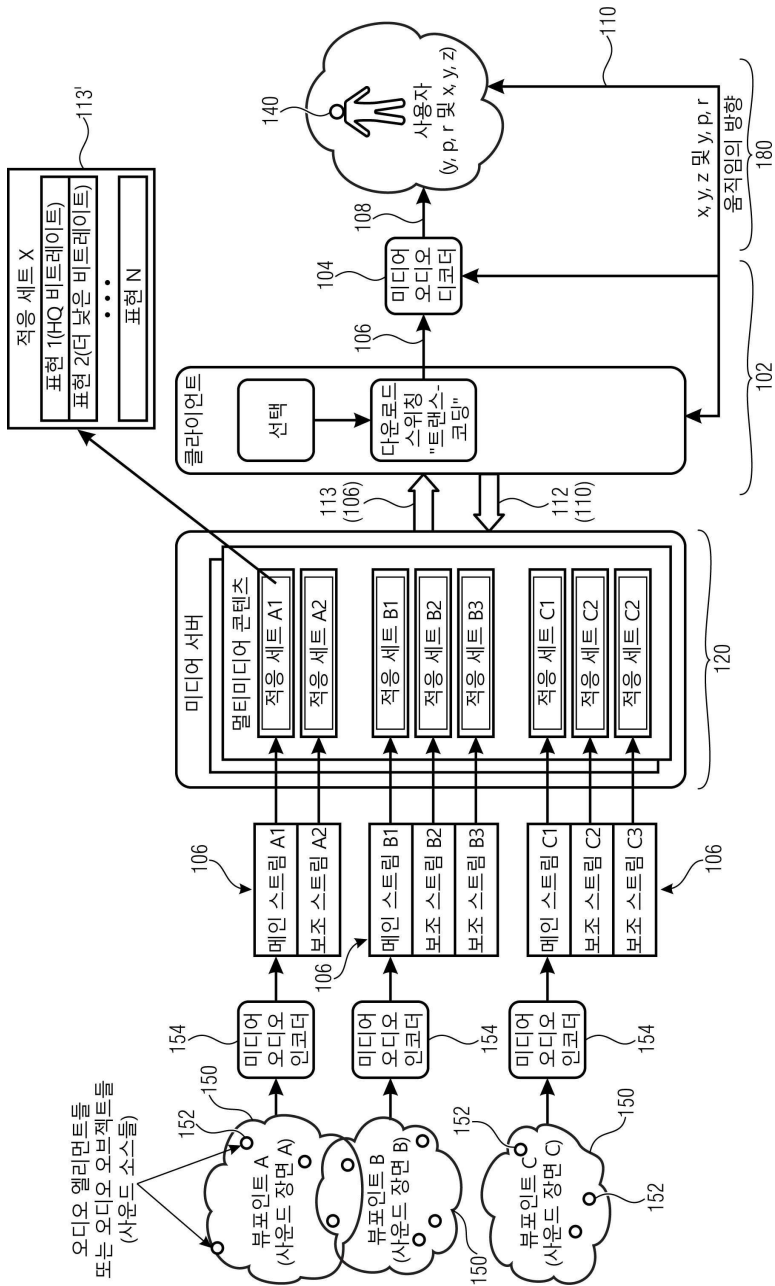
[0381] 본 명세서의 위에서, 가상 현실(VR), 증강 현실(AR), 혼합 현실(MR) 또는 360도 비디오 환경을 위해 오디오 및 비디오 스트림들을 클라이언트에 전달하기 위한 서버(120)가 참조되며, 비디오 및 오디오 스트림들은 미디어 소비 디바이스에서 재생될 것이고, 서버(120)는, 시각적 환경을 설명하기 위한 비디오 스트림들을 인코딩하기 위

한 인코더 및/또는 시각적 환경을 설명하기 위한 비디오 스트림들을 저장하기 위한 저장소를 포함하고, 시각적 환경은 오디오 장면에 연관되고; 서버는, 클라이언트에 전달될 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 인코딩하기 위한 인코더 및/또는 클라이언트에 전달될 복수의 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들을 저장하기 위한 저장소를 더 포함하며, 스트림들 및/또는 오디오 엘리먼트들 및/또는 적응 세트들은 적어도 하나의 오디오 장면에 연관되고, 서버는,

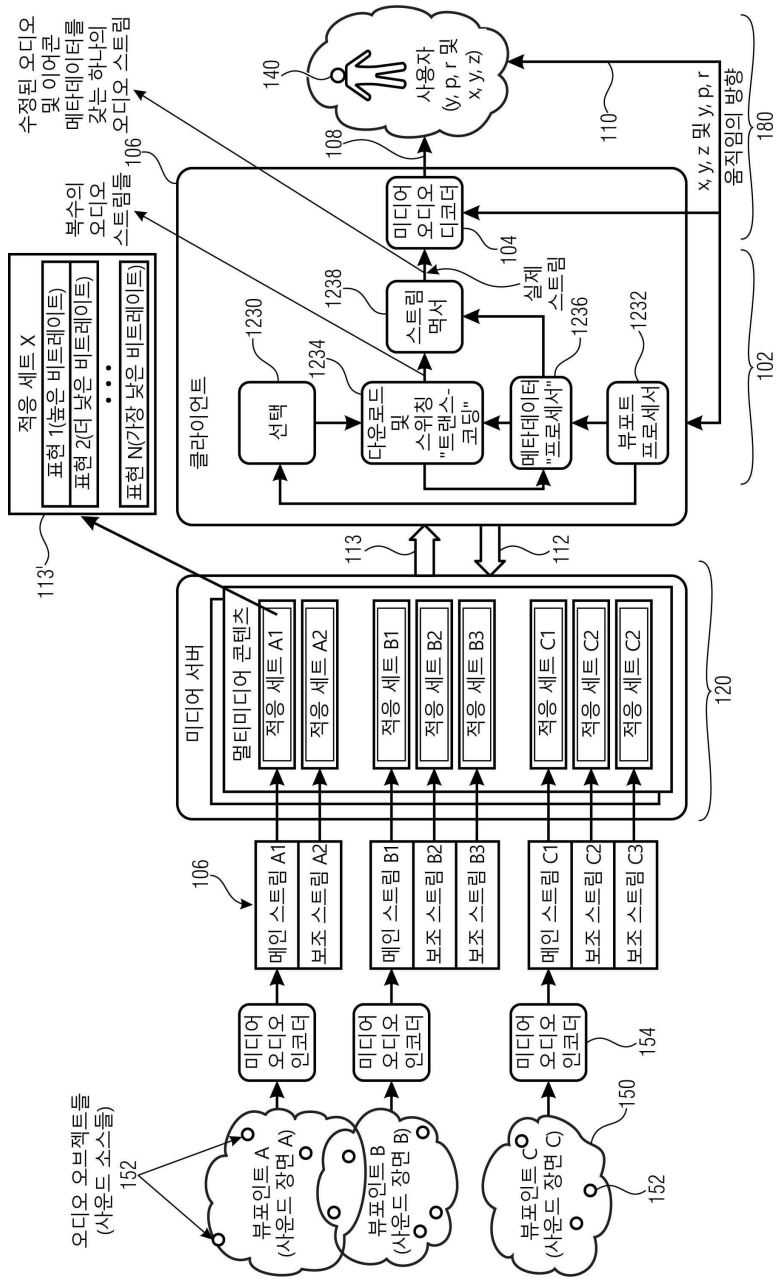
- [0382] 클라이언트로부터의 요청에 기초하여, 환경에 연관된 비디오 스트림을 선택 및 전달하고;
- [0383] 사용자의 현재 뷰포트 및/또는 머리 배향 및/또는 움직임 데이터 및/또는 상호작용 메타데이터 및/또는 가상 위치 데이터에, 그리고 환경에 연관된 오디오 장면에 적어도 연관된 클라이언트로부터의 요청에 기초하여 오디오 스트림 및/또는 오디오 엘리먼트 및/또는 적응 세트를 선택하며; 그리고
- [0384] 오디오 스트림을 클라이언트에 전달하도록 이루어진다.
- [0385] 추가적인 실시예들 및 변형들
- [0386] 특정한 구현 요건들에 의존하여, 예들은 하드웨어로 구현될 수 있다. 구현은, 개개의 방법이 수행되도록 프로그래밍가능 컴퓨터 시스템과 협력하는(또는 협력할 수 있는), 전자적으로 판독가능한 제어 신호들이 저장된 디지털 저장 매체, 예를 들어 플로피 디스크, 디지털 다목적 디스크(Digital Versatile Disc; DVD), 블루-레이 디스크, 콤팩트 디스크(Compact Disc; CD), 판독-전용 메모리(Read-only Memory; ROM), 프로그래밍가능 판독-전용 메모리(Programmable Read-only Memory; PROM), 소거가능 및 프로그래밍가능 판독-전용 메모리(Erasable and Programmable Read-only Memory; EPROM), 전기적으로 소거가능한 프로그래밍가능 판독-전용 메모리(Electrically Erasable Programmable Read-Only Memory; EEPROM) 또는 플래시 메모리를 사용하여 수행될 수 있다. 따라서, 디지털 저장 매체는 컴퓨터 판독가능할 수 있다.
- [0387] 일반적으로, 예들은 프로그램 명령들을 갖는 컴퓨터 프로그램 제품으로서 구현될 수 있으며, 프로그램 명령들은, 컴퓨터 프로그램 제품이 컴퓨터 상에서 구동되는 경우 방법들 중 하나를 수행하기 위해 동작된다. 프로그램 명령들은, 예를 들어, 머신 판독가능 매체 상에 저장될 수 있다.
- [0388] 다른 예들은, 머신 판독가능 캐리어 상에 저장되는, 본 명세서에 설명된 방법들 중 하나를 수행하기 위한 컴퓨터 프로그램을 포함한다. 따라서, 다시 말하면, 방법의 일 예는, 컴퓨터 프로그램이 컴퓨터 상에서 구동되는 경우, 본 명세서에 설명된 방법들 중 하나를 수행하기 위한 프로그램 명령들을 갖는 컴퓨터 프로그램이다.
- [0389] 따라서, 방법들의 추가적인 예는, 본 명세서에 설명된 방법들 중 하나를 수행하기 위한 컴퓨터 프로그램(데이터 캐리어 매체 상에 기록되어 있음)을 포함하는 데이터 캐리어 매체(또는 디지털 저장 매체, 또는 컴퓨터-판독가능 매체)이다. 데이터 캐리어 매체, 디지털 저장 매체 또는 기록 매체는 무형의 그리고 일시적인 신호보다는 유형의 그리고/또는 비-일시적이다.
- [0390] 추가적인 예는, 본 명세서에 설명된 방법들 중 하나를 수행하는 프로세싱 유닛, 예를 들어, 컴퓨터, 또는 프로그래밍가능 로직 디바이스를 포함한다.
- [0391] 추가적인 예는, 본 명세서에 설명된 방법들 중 하나를 수행하기 위한 컴퓨터 프로그램이 설치된 컴퓨터를 포함한다.
- [0392] 추가적인 예는, 본 명세서에 설명된 방법들 중 하나를 수행하기 위한 컴퓨터 프로그램을 (예를 들어, 전자적으로 또는 광학적으로) 수신기에 전달하는 장치 또는 시스템을 포함한다. 수신기는, 예를 들어, 컴퓨터, 모바일 디바이스, 메모리 디바이스 등일 수 있다. 장치 또는 시스템은, 예를 들어, 컴퓨터 프로그램을 수신기에 전달하기 위한 파일 서버를 포함할 수 있다.
- [0393] 일부 예들에서, 프로그래밍가능 로직 디바이스(예를 들어, 필드 프로그래밍가능 게이트 어레이)는, 본 명세서에 설명된 방법들의 기능들 중 일부 또는 모두를 수행하기 위해 사용될 수 있다. 일부 예들에서, 필드 프로그래밍가능 게이트 어레이는, 본 명세서에 설명된 방법들 중 하나를 수행하기 위해 마이크로프로세서와 협력할 수 있다. 일반적으로, 방법들은 임의의 적절한 하드웨어 장치에 의해 수행될 수 있다.
- [0394] 위에서 설명된 예들은 위에서 논의된 원리들에 대해 예시적이다. 본 명세서에 설명된 배열들 및 세부사항들의 수정들 및 변경들이 명백할 것임을 이해한다. 따라서, 본 명세서의 예들의 설명 및 해설에 의해 제시된 특정 세부사항들이 아니라 임박한 특허 청구항들의 범위에 의해 제한되는 것이 의도이다.

도면

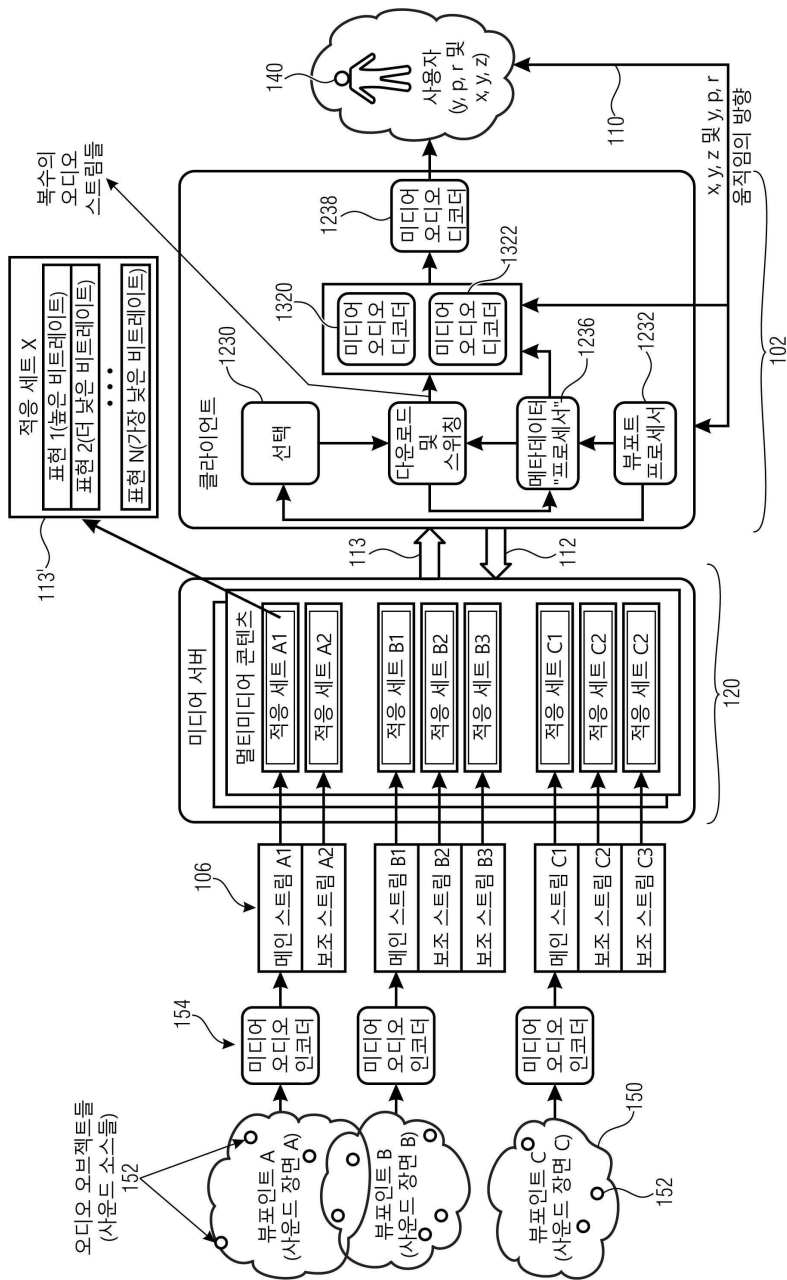
도면1a



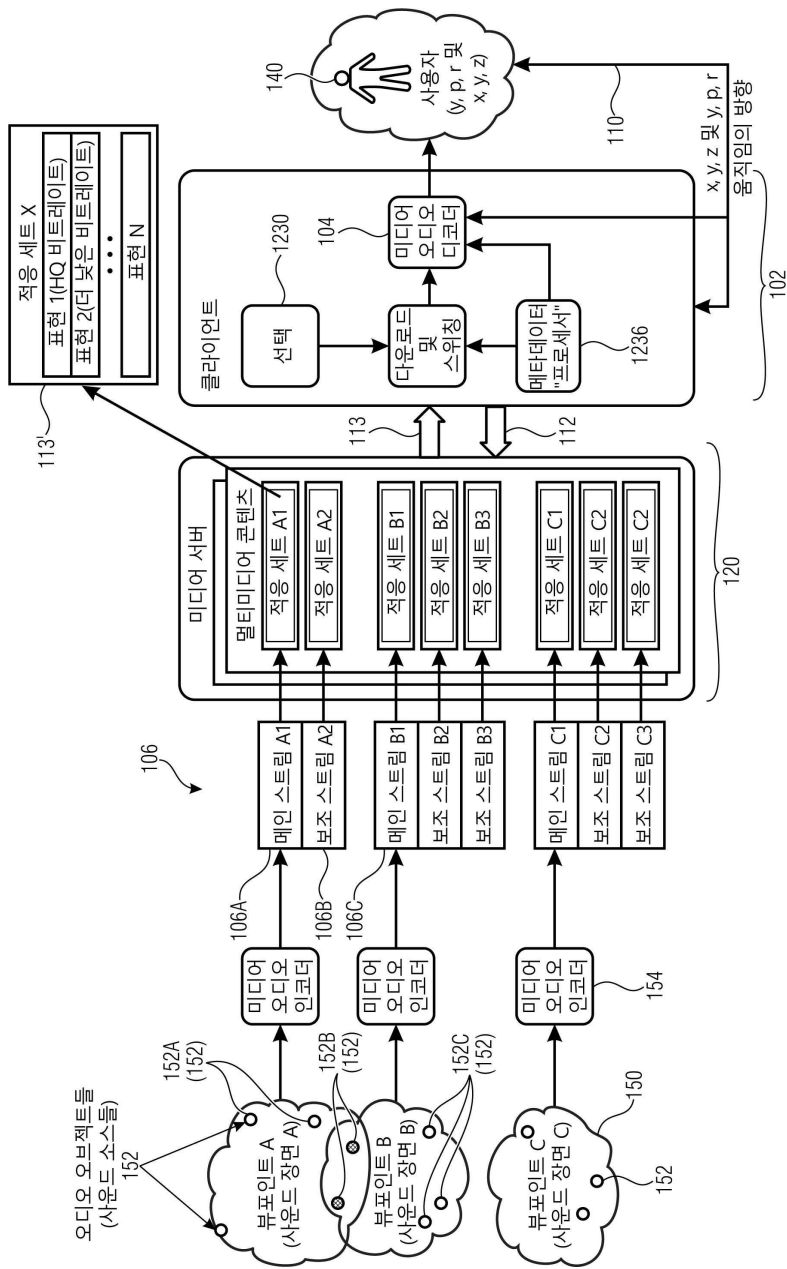
도면 1b



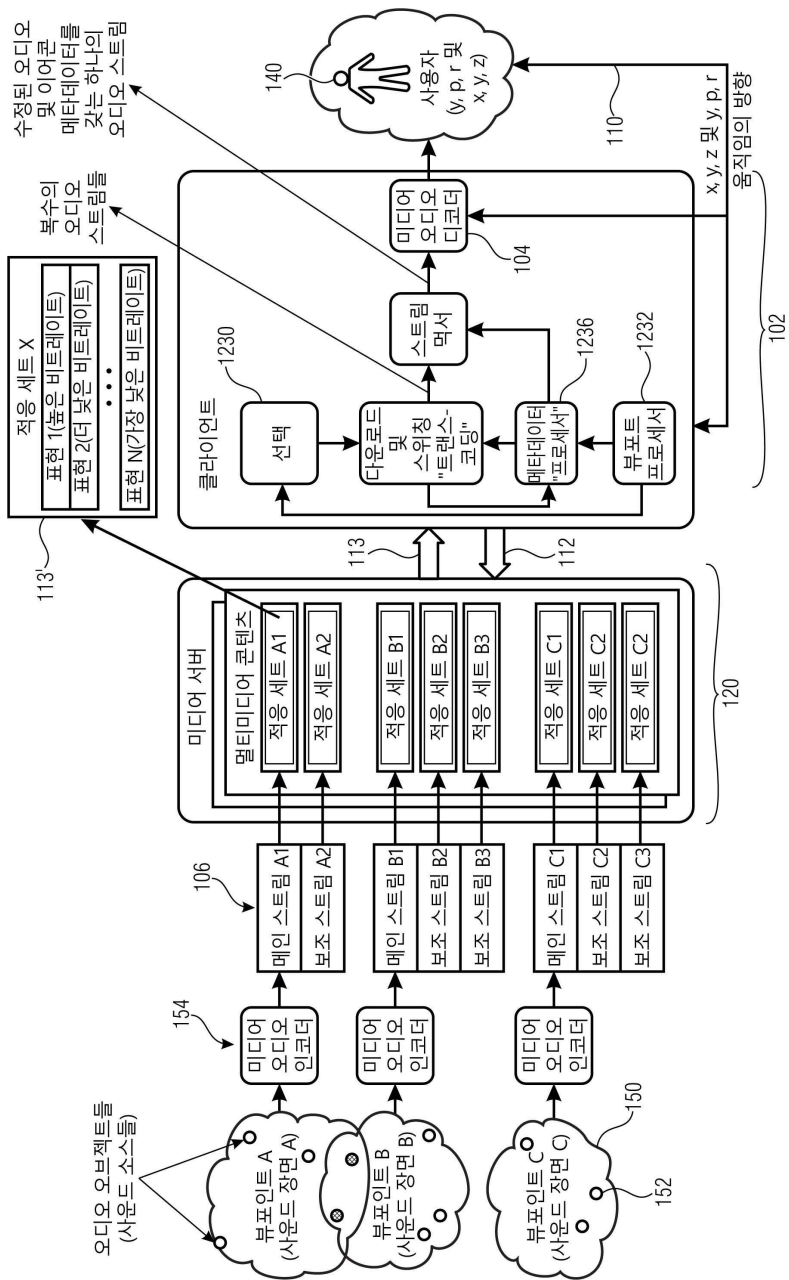
도면1c



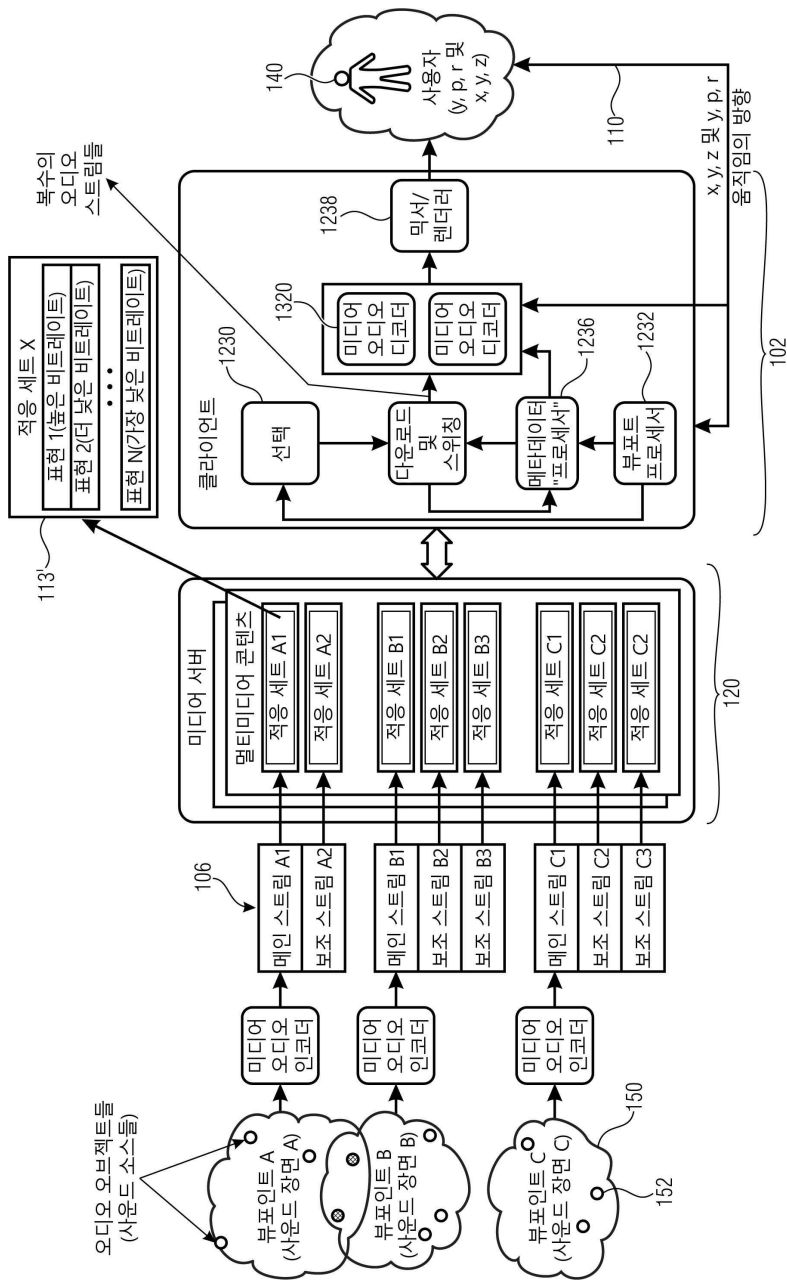
도면1d



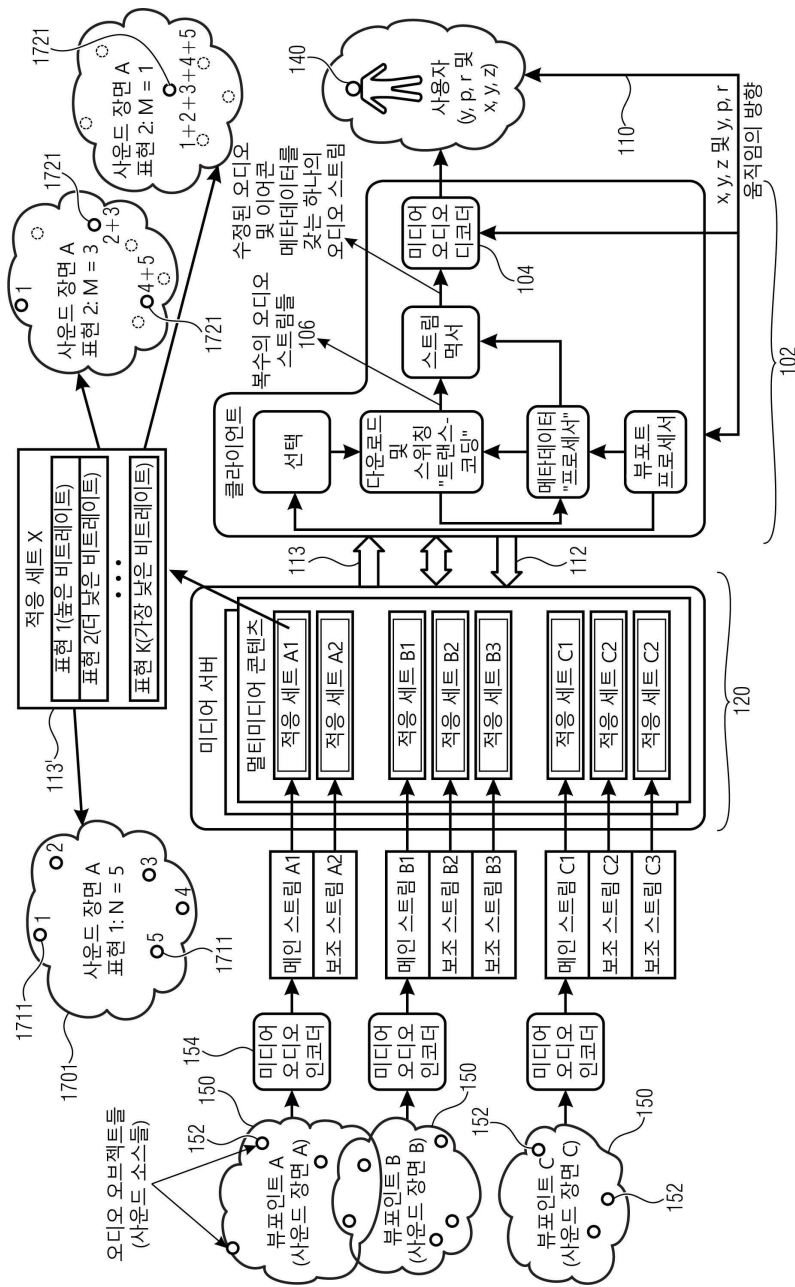
도면 1e



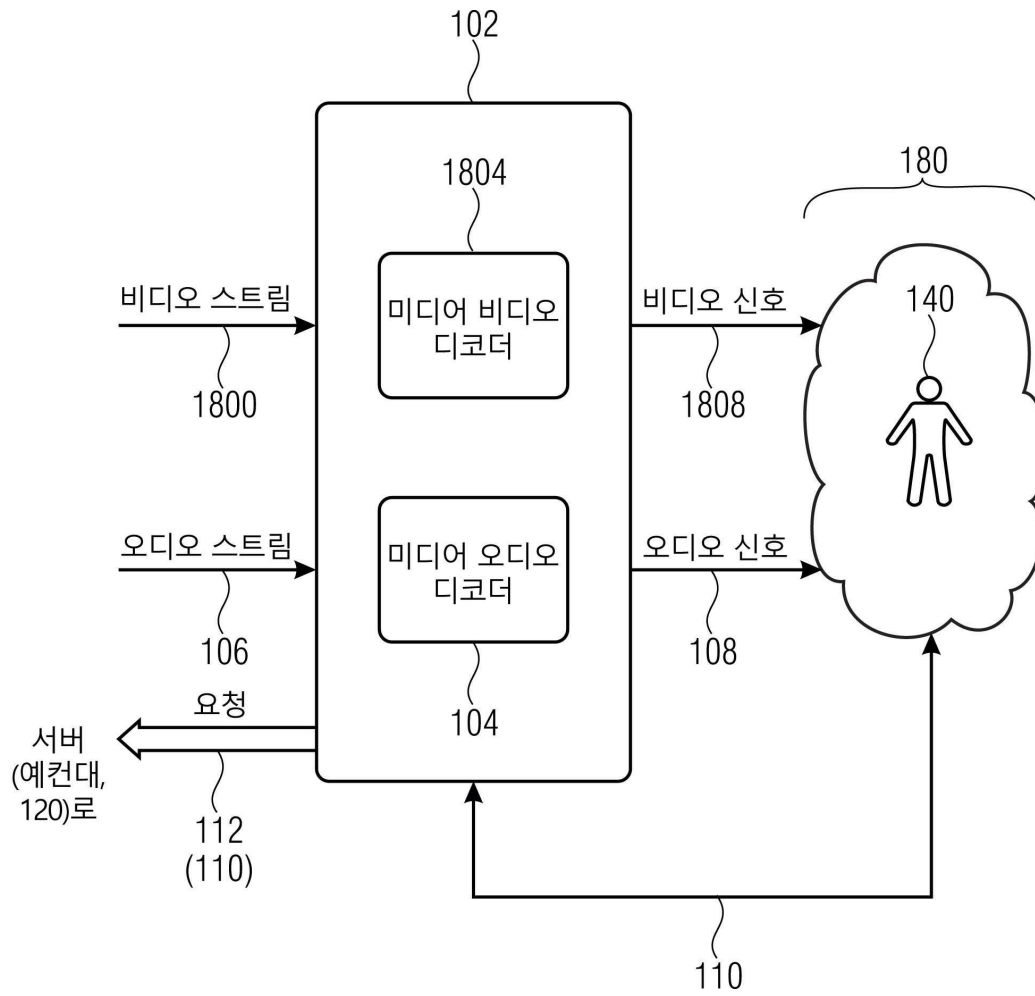
도면 1f



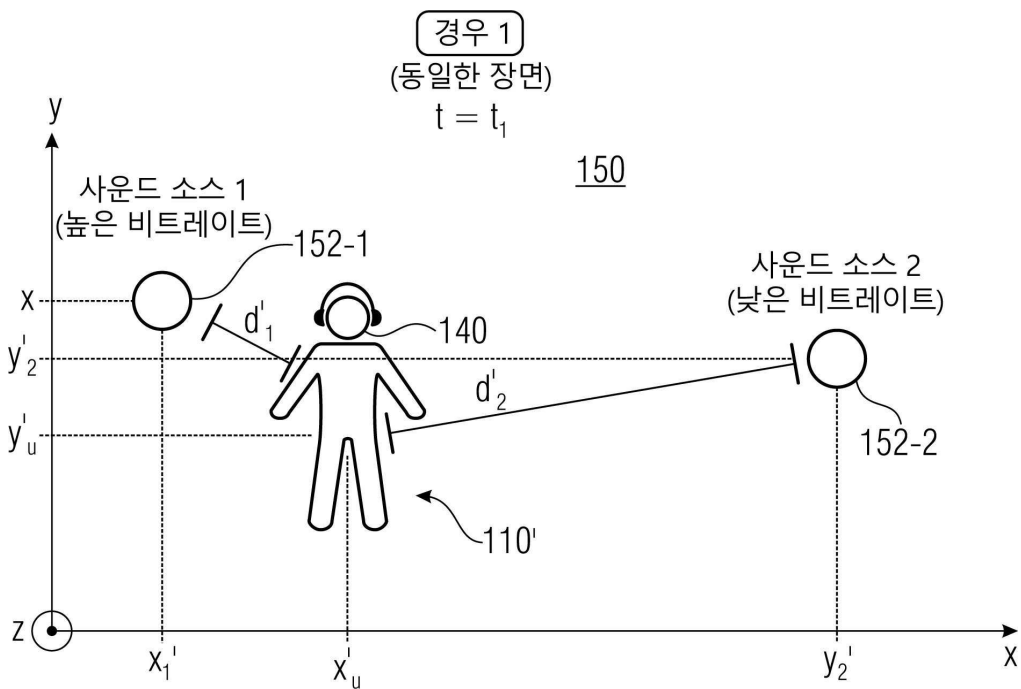
도면 1g



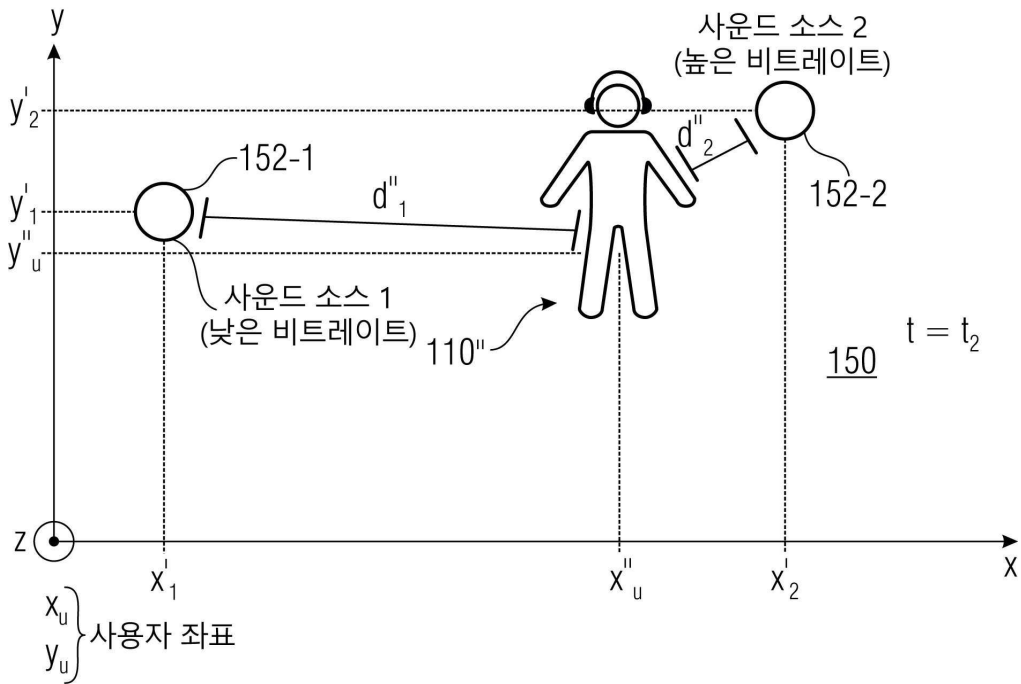
도면1h



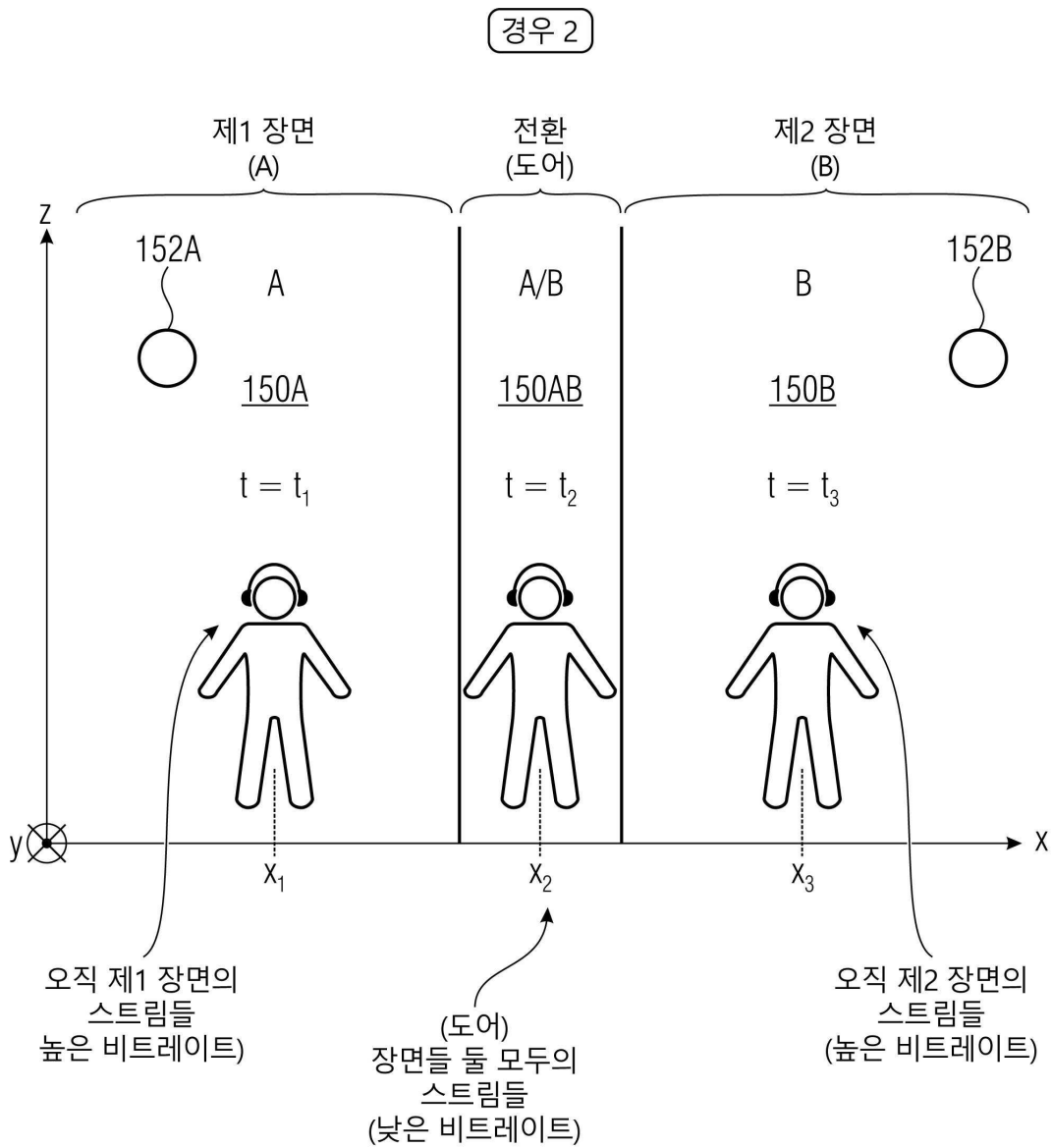
도면2a



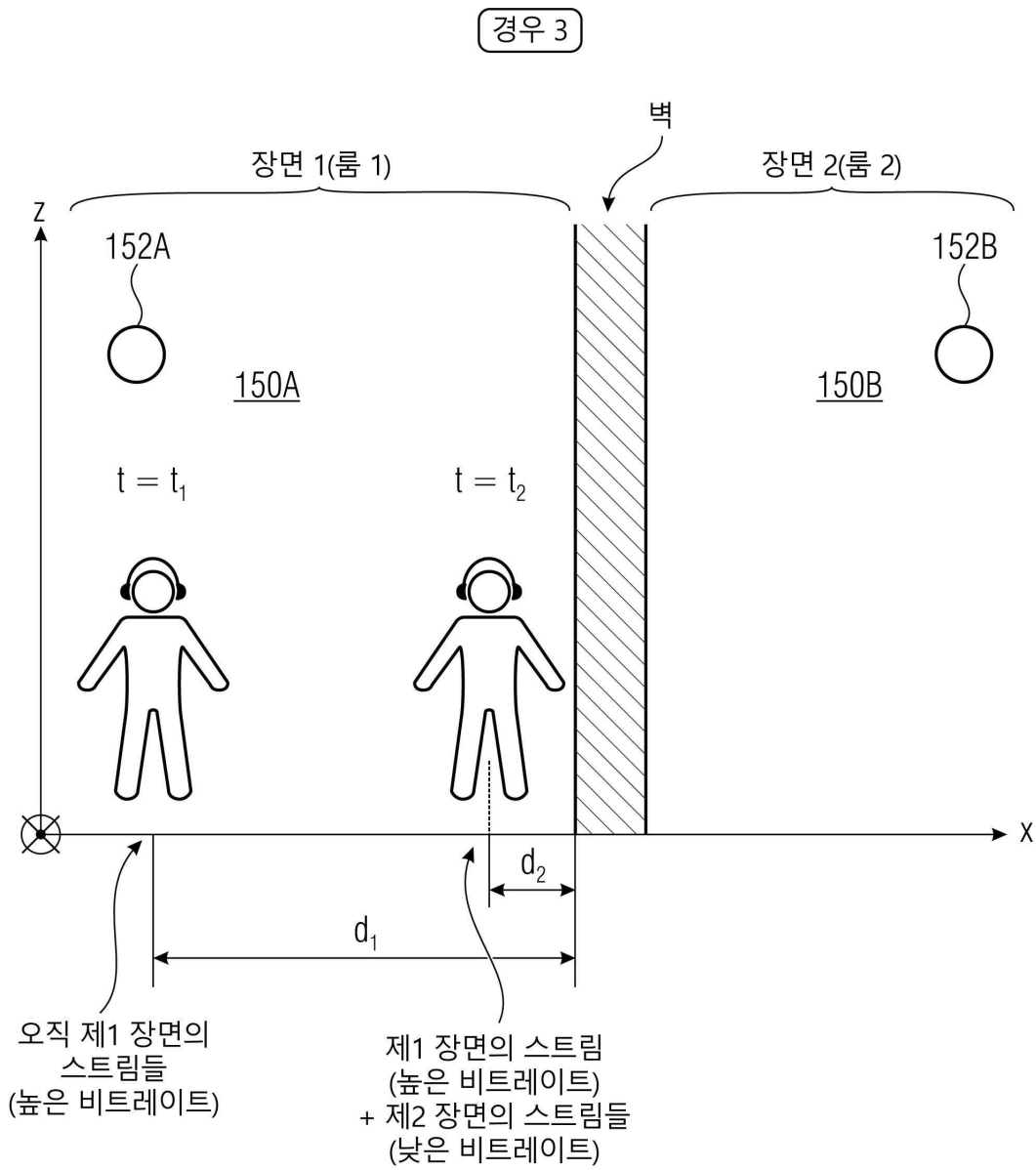
도면 2b



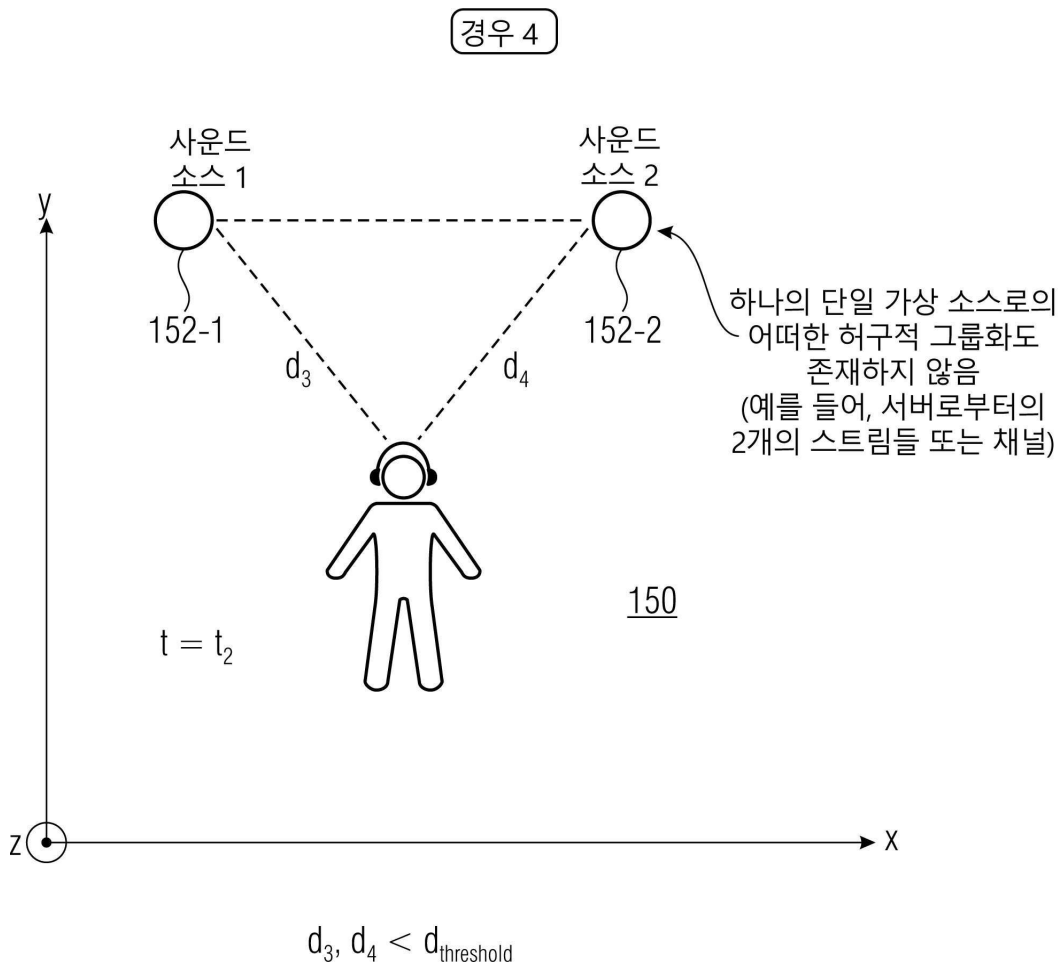
도면3



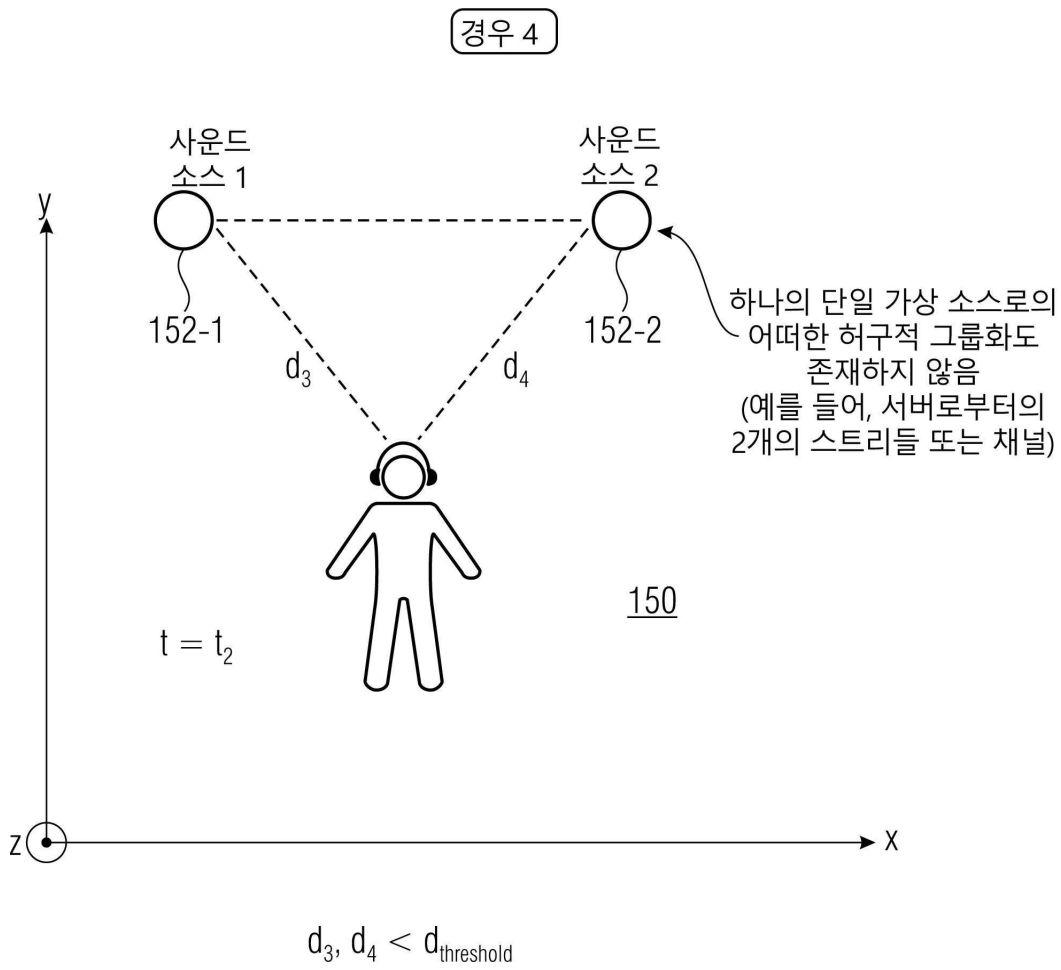
도면4



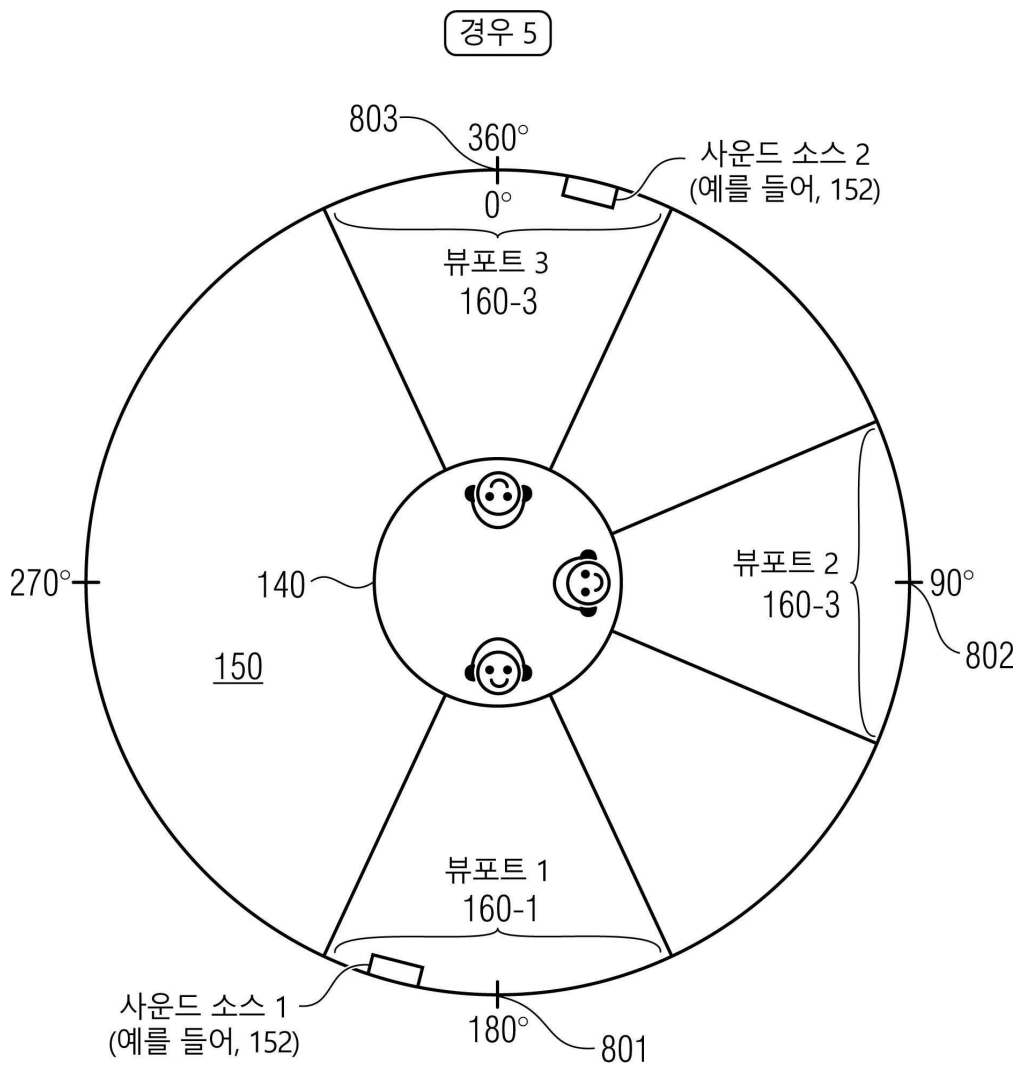
도면5a



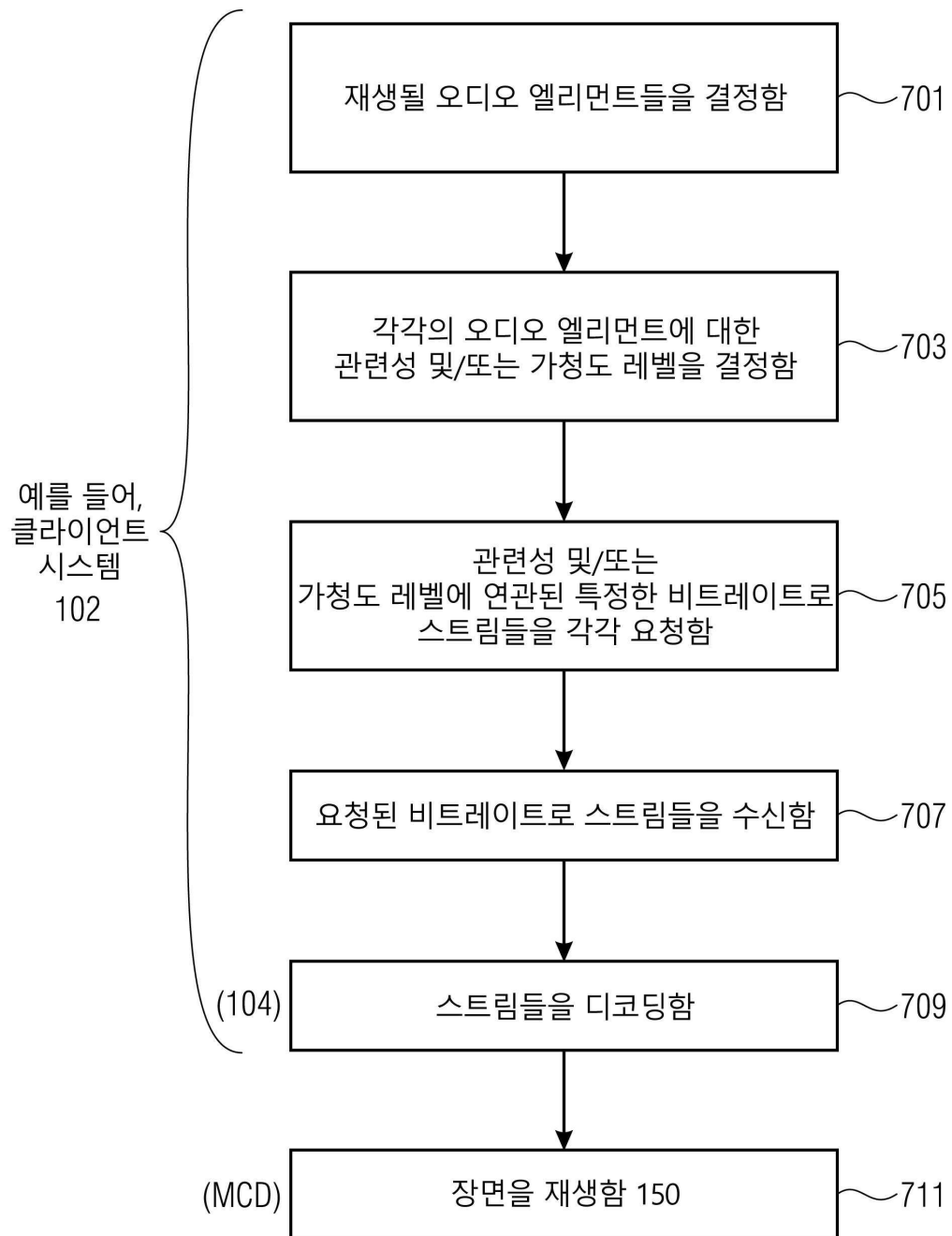
도면 5b



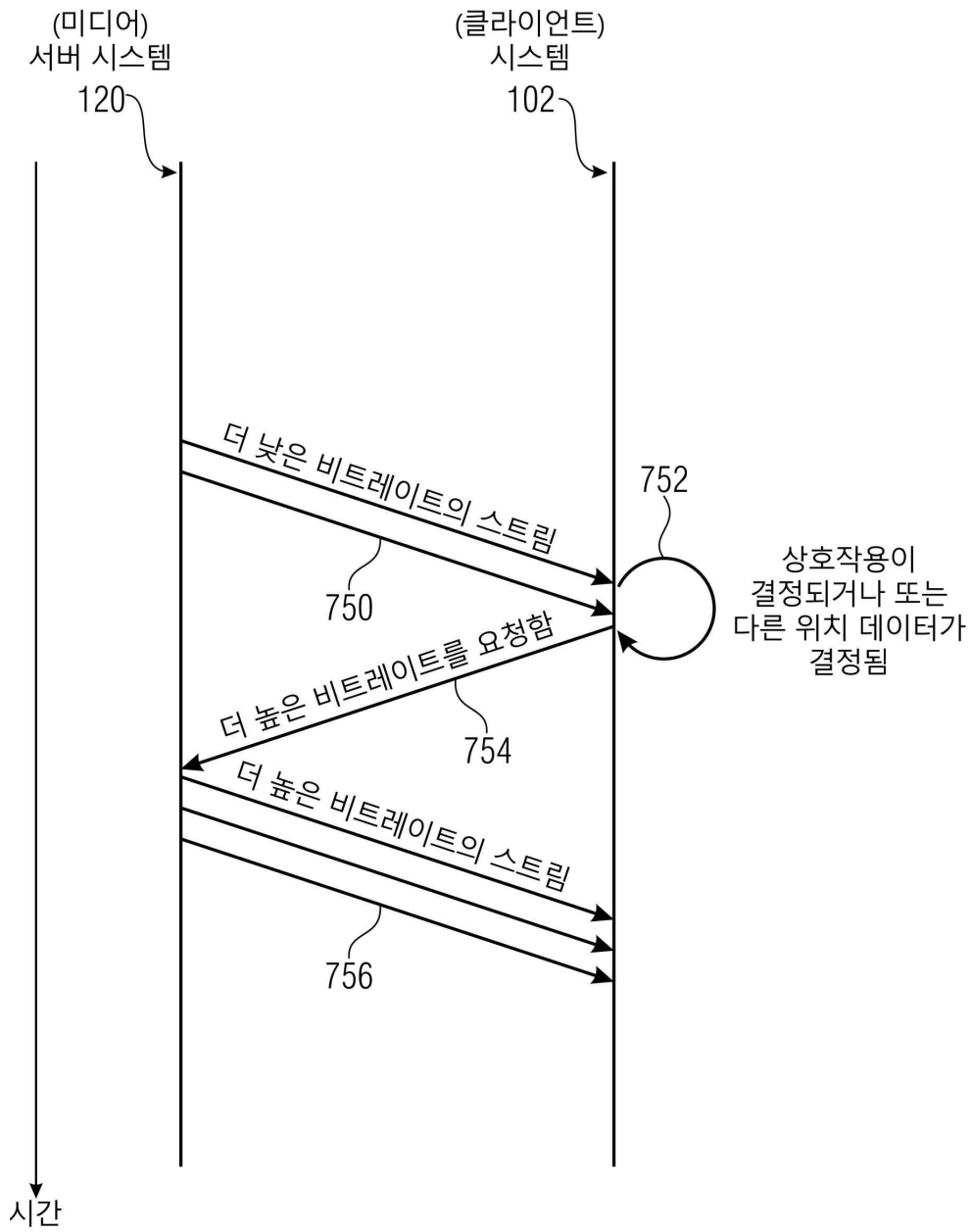
도면6



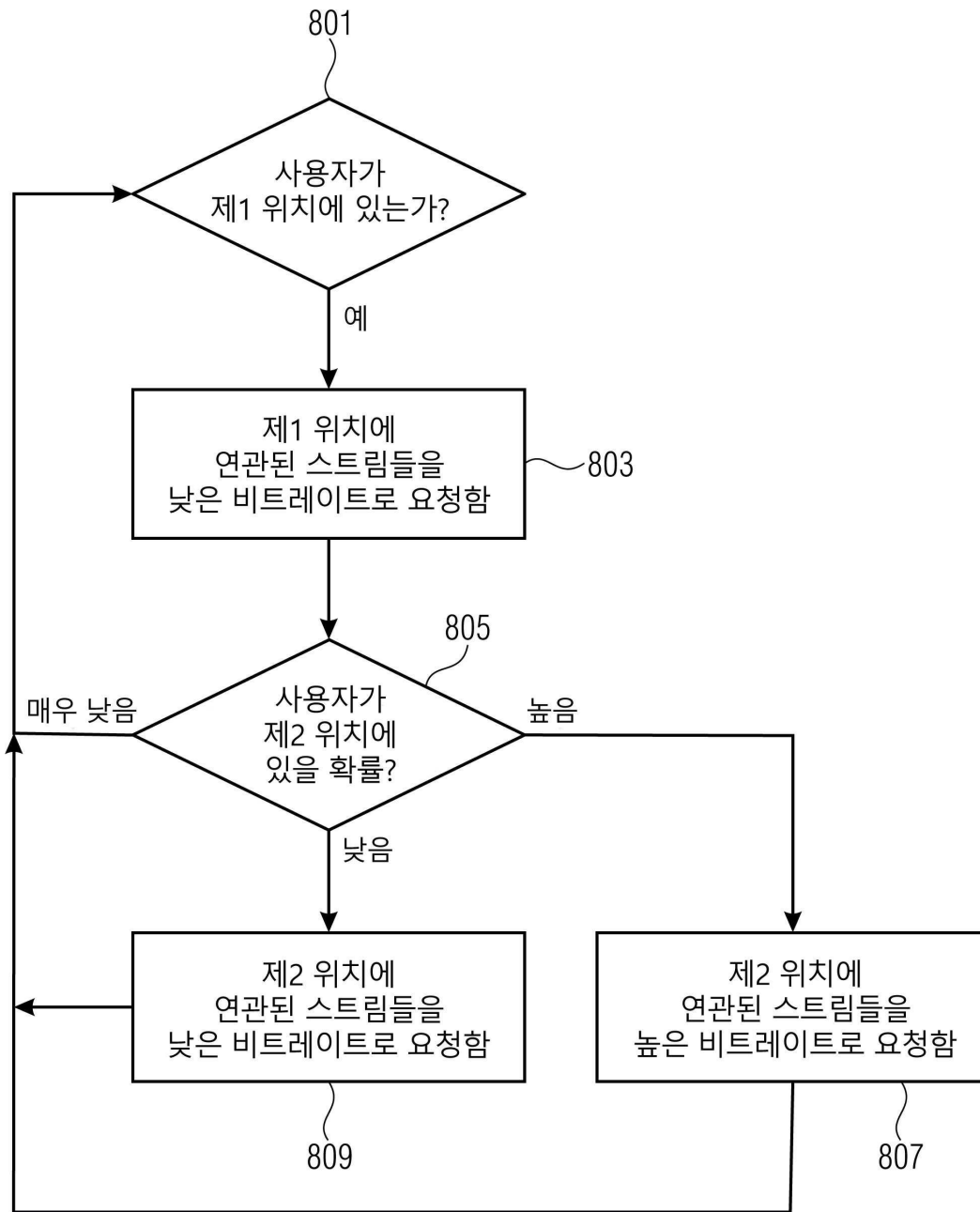
도면7a



도면 7b



도면 8a



도면 8b

