



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년07월07일

(11) 등록번호 10-2419065

(24) 등록일자 2022년07월05일

- (51) 국제특허분류(Int. Cl.)  
*G06F 3/0481* (2022.01) *A63F 13/424* (2014.01)  
*A63F 13/5255* (2014.01) *A63F 13/5372* (2014.01)  
*A63F 13/54* (2014.01) *A63F 13/65* (2014.01)  
*G02B 27/00* (2020.01) *G02B 27/01* (2006.01)  
*G06F 3/01* (2006.01) *G06F 3/0484* (2022.01)  
*G06F 3/16* (2018.01)
- (52) CPC특허분류  
*G06F 3/04815* (2022.01)  
*A63F 13/424* (2015.01)
- (21) 출원번호 10-2021-7031142(분할)  
(22) 출원일자(국제) 2018년02월27일  
심사청구일자 2021년09월28일  
(85) 번역문제출일자 2021년09월28일  
(65) 공개번호 10-2021-0119589  
(43) 공개일자 2021년10월05일  
(62) 원출원 특허 10-2019-7028175  
원출원일자(국제) 2018년02월27일  
심사청구일자 2021년02월26일  
(86) 국제출원번호 PCT/US2018/020020  
(87) 국제공개번호 WO 2018/160593  
국제공개일자 2018년09월07일  
(30) 우선권주장  
62/464,757 2017년02월28일 미국(US)  
(56) 선행기술조사문헌  
JP2004267433 A\*  
KR1020150118855 A\*  
KR1020160018436 A\*  
KR1020160079788 A\*  
\*는 심사관에 의하여 인용된 문헌

- (73) 특허권자  
**매직 립, 인코포레이티드**  
미국 플로리다 플랜타타운 웨스트 선라이즈 블러바드 7500 (우: 33322)
- (72) 발명자  
**후앙, 지치앙**  
미국 33332 플로리다 웨스턴 노스 허비스커스 스트리트19226
- (74) 대리인  
**특허법인 남앤남**

전체 청구항 수 : 총 20 항

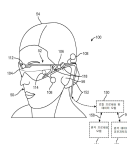
심사관 : 장재우

(54) 발명의 명칭 **혼합 현실 디바이스에서의 가상 및 실제 객체 레코딩**

### (57) 요약

최종 사용자에게 의한 사용을 위한 가상 이미지 생성 시스템은, 메모리, 디스플레이 서브시스템, 최종 사용자로부터 입력을 수신하고, 최종 사용자 입력에 대한 응답으로 적어도 하나의 객체를 지속적으로 선택하도록 구성된 객체 선택 디바이스; 및 3-차원 장면의 복수의 이미지 프레임들을 렌더링하고, 이미지 프레임들을 디스플레이 서브시스템에 전달하고, 적어도 하나의 선택된 객체로부터 유래한 오디오 데이터를 생성하고, 오디오 데이터를 메모리 내에 저장하도록 구성된 제어 서브시스템을 포함한다.

### 대표도



(52) CPC특허분류

*A63F 13/5255* (2015.01)  
*A63F 13/5372* (2015.01)  
*A63F 13/54* (2015.01)  
*A63F 13/65* (2015.01)  
*G02B 27/0093* (2013.01)  
*G02B 27/0172* (2013.01)  
*G06F 3/012* (2013.01)  
*G06F 3/017* (2013.01)  
*G06F 3/04842* (2022.01)

---

## 명세서

### 청구범위

#### 청구항 1

증강 현실 환경에서 공간화된 오디오를 갖는 오디오-시각 효과들을 재생하기 위한 방법으로서,

제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계 — 상기 증강 현실 경험은 실제 객체들과 가상 객체들을 동시에 보는 것을 포함함 —; 및

제2 증강 현실 환경에서 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 단계

를 포함하고,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계는, 상기 제1 증강 현실 환경에서 캡처된 오디오 데이터 및 상기 오디오 데이터와 연관된 제1 메타데이터를 식별하는 단계를 포함하고, 상기 오디오 데이터는 제1 환경 내의 제1 사용자 위치에 있는 상기 제1 사용자에게 의해 상기 제1 증강 현실 환경 내의 객체로부터 나오는 것처럼 지각(perceive)될 수 있고, 상기 제1 메타데이터는 상기 오디오 데이터가 레코딩된 상기 제1 증강 현실 환경의 적어도 일부 및 상기 제1 사용자에게 관련되고, 상기 제1 증강 현실 환경은 증강 현실 시스템에 의해 제1 실세계 환경 부분으로부터 생성되고,

상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 단계는,

적어도, 제2 메타데이터에 적어도 부분적으로 기초하여 상기 제1 메타데이터를 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하여 제1 로컬화된 메타데이터를 생성함으로써, 상기 오디오 데이터를 공간화된 오디오로 렌더링하는 단계 — 상기 제2 메타데이터는 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대한 가상 이미지 생성 시스템에 의해 검출된 위치 또는 배향 특성에 관련됨 —; 및

상기 제1 메타데이터 및 상기 제2 메타데이터에 적어도 부분적으로 기초하여 적어도 상기 객체에 관련된 적어도 하나의 가상 객체 및 상기 공간화된 오디오를 동시에 사용자에게 제공(present)하는 단계

를 포함하는,

방법.

#### 청구항 2

제1항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계는,

상기 제1 증강 현실 환경의 상기 제1 실세계 환경 부분의 적어도 일부의 실제 시각 데이터를 캡처하는 단계 — 상기 실제 시각 데이터는 상기 제1 실세계 환경에 대한 실제 이미지 데이터 또는 실제 비디오 데이터를 포함하고, 상기 실제 이미지 데이터는 실제 이미지 콘텐츠 데이터 및 상기 실제 이미지 콘텐츠 데이터와 연관된 이미지 메타데이터를 포함하고, 상기 실제 비디오 데이터는 실제 비디오 콘텐츠 데이터 및 상기 실제 비디오 콘텐츠 데이터와 연관된 실제 비디오 메타데이터를 포함함 —; 및

상기 실제 시각 데이터와 가상 시각 데이터를, 혼합된 시각 데이터로 혼합하는 단계 — 상기 가상 시각 데이터는 상기 제1 증강 현실 환경에서의 제1 가상 환경 부분에 대한 가상 이미지 데이터 또는 가상 비디오 데이터를 포함하고, 상기 가상 시각 데이터는 상기 증강 현실 시스템에 의해 생성됨 —

를 더 포함하는,

방법.

#### 청구항 3

제1항에 있어서,

상기 오디오 데이터를 식별하는 것은, 상기 증강 현실 시스템에서 수신된 사용자 입력에 응답하여, 상기 증강 현실 시스템의 객체 선택 디바이스에서, 상기 제1 증강 현실 환경 내의 복수의 객체들 중 상기 객체를 선택하는 것을 포함하고,

상기 복수의 객체들은 적어도 하나의 실제 객체 및 적어도 하나의 가상 객체를 포함하는, 방법.

#### 청구항 4

제3항에 있어서,

상기 오디오 데이터를 식별하는 것은, 별개의 사용자 입력에 응답하여, 상기 증강 현실 시스템의 상기 객체 선택 디바이스에서 선택되었던 제1 객체를 선택 해제하는 것을 더 포함하고,

상기 객체 선택 디바이스는, 상기 사용자에게 디스플레이되는 커서를 제어하는 물리적 제어기, 오디오 데이터 캡처 디바이스, 하나 이상의 음성 커맨드들을 해석하는 음성 해석 모듈, 제스처들을 추적하는 제1 시각 데이터 검출 디바이스, 상기 사용자의 눈의 위치 또는 배향을 추적하는 제2 시각 데이터 검출 디바이스, 또는 이들의 조합 중 적어도 하나를 포함하는,

방법.

#### 청구항 5

제1항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계는,

상기 제1 증강 현실 환경 내의 다수의 객체들로부터 상기 객체를 선택하는 단계;

상기 증강 현실 시스템을 이용하여 상기 객체의 위치 또는 배향을 추적하는 단계;

상기 객체의 위치 또는 배향에 적어도 부분적으로 기초하여 상기 객체로부터 나오는 소리를 우선적으로 감지하는 단계

를 더 포함하고,

상기 오디오 데이터는, 상기 제1 증강 현실 환경에서 제1 방향, 제1 객체 위치, 또는 제1 객체 자세에서 다른 객체보다 상기 객체에 대해 우선적으로 레코딩되는,

방법.

#### 청구항 6

제5항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계는,

상기 증강 현실 시스템의 위치 및 포즈 추적 디바이스에서, 상기 제1 증강 현실 환경 내의 상기 객체에 대한 상기 제1 사용자 또는 상기 제1 사용자의 부분의 위치 및 포즈 또는 배향을 지속적으로 그리고 우선적으로 검출하는 단계;

상기 객체에 대한 상기 제1 사용자 또는 상기 제1 사용자의 일부의 위치 및 포즈 또는 배향을 지속적으로 그리고 우선적으로 검출한 결과에 적어도 부분적으로 기초하여, 상기 제1 사용자에 관한 상기 제1 메타데이터를 업데이트된 제1 메타데이터로 업데이트하는 단계; 및

상기 업데이트된 제1 메타데이터에 적어도 부분적으로 기초하여 상기 제1 로컬화 메타데이터를 업데이트하는 단계

를 더 포함하는,

방법.

## 청구항 7

제6항에 있어서,

상기 공간화된 오디오는, 상기 제2 증강 현실 환경 내의 제1 실제 객체 위치에 있는 제1 실제 객체에 의해 반사되거나, 차폐되거나 또는 가려진 것처럼 상기 제2 증강 현실 환경 내의 상기 제1 사용자 또는 제2 사용자에게 의해 지각되는,

방법.

## 청구항 8

제1항에 있어서,

상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 단계는, 상기 공간화된 오디오가 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 의해, 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 대해 상대적인 제1 가상 객체 위치에 있는 제1 가상 객체로부터 나오는 것처럼 지각 되도록, 상기 공간화된 오디오를 상기 제2 증강 현실 환경 내의 상기 제1 가상 객체 위치에 있는 상기 제1 가상 객체에 대해 렌더링하는 단계를 더 포함하고,

상기 오디오 데이터는 상기 제1 증강 현실 환경 내의 상기 객체로부터 나오는 것처럼 캡처되었고, 상기 객체는 상기 증강 현실 시스템에 의해 상기 제2 증강 현실 환경 내의 상기 제1 가상 객체로서 렌더링되는,

방법.

## 청구항 9

제1항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 단계는,

적어도 오디오 프로세싱 모듈을 이용하여 상기 제1 환경에서 상기 오디오 데이터를 캡처하는 단계 — 상기 오디오 프로세싱 모듈은 마이크로폰, 지연 엘리먼트, 및 이득 엘리먼트를 포함하고, 상기 증강 현실 시스템에 저장된 극 패턴(polar pattern)에 적어도 부분적으로 기초하여 동작함 —;

상기 증강 현실 시스템의 이미지 캡처 디바이스를 이용하여 이미지 또는 비디오 콘텐츠 데이터를 캡처하는 단계;

상기 증강 현실 시스템의 추적 디바이스 또는 위치결정 디바이스를 이용하여 상기 이미지 또는 비디오 콘텐츠 데이터와 연관된 이미지 또는 비디오 데이터를 캡처하는 단계; 및

상기 오디오 데이터에 관련된 오디오 메타데이터를 결정하는 단계 — 상기 오디오 메타데이터를 결정하는 단계는 적어도 상기 오디오 데이터를 상기 이미지 또는 비디오 콘텐츠 데이터 또는 상기 이미지 또는 비디오 메타데이터와 연관시킴으로써 수행됨 —

를 더 포함하는,

방법.

## 청구항 10

제1항에 있어서,

상기 제1 메타데이터를, 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 단계는,

상기 제1 환경에서 상기 제1 사용자의 제1 사용자 위치 또는 제1 포즈에 대한 상기 객체의 위치 또는 배향에 관련된 상대 위치 또는 배향 데이터를 식별하는 단계;

상기 제1 사용자 위치에 있는 상기 제1 사용자에게 의해 지각되는 상기 오디오 데이터를 캡처하는 단계 — 상기 오디오 데이터를 캡처하는 단계는 적어도, 상기 제1 환경에서 상기 증강 현실 시스템의 마이크로폰 엘리먼트, 지연 엘리먼트, 이득 엘리먼트, 및 극 방향성 패턴을 이용함으로써 수행됨 —;

상기 상대 위치 또는 배향 데이터를 상기 오디오 데이터와 연관시키는 단계;

상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에서의 위치 또는 배향 특성을 식별하는 단계; 및

상기 제1 메타데이터를 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 단계  
 — 상기 제1 메타데이터를 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 단계는 적어도, 상기 제2 사용자 위치에서의 상기 위치 또는 배향 특성에 적어도 부분적으로 기초하여 상기 상대 위치 또는 배향 데이터를 상기 제1 사용자 또는 제2 사용자의 머리에 적용함으로써 수행됨 —

를 더 포함하는,

방법.

## 청구항 11

증강 현실 시스템으로서,

오디오 신호 프로세서를 포함하는 오디오 프로세싱 모듈;

그래픽 프로세싱 유닛을 포함하고 상기 오디오 프로세싱 모듈에 커플링되어, 제1 사용자에게 오디오-시각 효과들을 생성하기 위한 이미지 프로세싱 모듈;

적어도 마이크로프로세서를 포함하고 상기 오디오 프로세싱 모듈 및 상기 이미지 프로세싱 모듈에 커플링되어, 상기 제1 사용자에게 가상 콘텐츠를 제공하기 위한 가상 이미지 생성 모듈

을 포함하고,

상기 증강 현실 시스템은, 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 동작 및 제2 증강 현실 환경에서 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 동작을 포함하는 동작들의 세트를 수행하도록 구성되고 — 상기 증강 현실 경험은 실제 객체들과 가상 객체들을 동시에 보는 것을 포함함 —,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 동작은, 상기 제1 증강 현실 환경에서 캡처된 오디오 데이터 및 상기 오디오 데이터와 연관된 제1 메타데이터를 식별하는 동작을 포함하고, 상기 오디오 데이터는 제1 환경 내의 제1 사용자 위치에 있는 상기 제1 사용자에게 의해 상기 제1 증강 현실 환경 내의 객체로부터 나오는 것처럼 지각될 수 있고, 상기 제1 메타데이터는 상기 오디오 데이터가 레코딩된 상기 제1 증강 현실 환경의 적어도 일부 및 상기 제1 사용자에게 관련되고, 상기 제1 증강 현실 환경은 상기 증강 현실 시스템에 의해 제1 실세계 환경 부분으로부터 생성되고,

상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 동작은,

적어도, 제2 메타데이터에 적어도 부분적으로 기초하여 상기 제1 메타데이터를 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하여 제1 로컬화된 메타데이터를 생성함으로써, 상기 오디오 데이터를 공간화된 오디오로 렌더링하는 동작 — 상기 제2 메타데이터는 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대한 가상 이미지 생성 시스템에 의해 검출된 위치 또는 배향 특성에 관련됨 —;

상기 제1 메타데이터 및 상기 제2 메타데이터에 적어도 부분적으로 기초하여 적어도 상기 객체에 관련된 적어도 하나의 가상 객체 및 상기 공간화된 오디오를 동시에 상기 사용자에게 제공하는 동작

을 포함하는,

증강 현실 시스템.

## 청구항 12

제11항에 있어서,

상기 증강 현실 시스템은 이미지 또는 비디오 캡처 디바이스를 더 포함하고,

상기 동작들의 세트는,

상기 이미지 또는 비디오 캡처 디바이스에서, 상기 제1 증강 현실 환경의 상기 제1 실세계 환경 부분의 적어도

일부의 실제 시각 데이터를 캡처하는 동작 - 상기 실제 시각 데이터는 상기 제1 실세계 환경에 대한 실제 이미지 데이터 또는 실제 비디오 데이터를 포함하고, 상기 실제 이미지 데이터는 실제 이미지 콘텐츠 데이터 및 상기 실제 이미지 콘텐츠 데이터와 연관된 실제 이미지 메타데이터를 포함하고, 상기 실제 비디오 데이터는 실제 비디오 콘텐츠 데이터 및 상기 실제 비디오 콘텐츠 데이터와 연관된 실제 비디오 메타데이터를 포함함 -; 및

상기 이미지 프로세싱 모듈에서, 상기 실제 시각 데이터와 가상 시각 데이터를, 혼합된 시각 데이터로 혼합하는 동작 - 상기 가상 시각 데이터는 상기 제1 증강 현실 환경에서의 제1 가상 환경 부분에 대한 가상 이미지 데이터 또는 가상 비디오 데이터를 포함하고, 상기 가상 시각 데이터는 상기 증강 현실 시스템에 의해 생성됨 -

을 더 포함하는,

증강 현실 시스템.

### 청구항 13

제11항에 있어서,

상기 증강 현실 시스템은 사용자들로부터 입력을 수신하는 사용자 입력 모듈을 더 포함하고,

상기 오디오 데이터를 식별하는 것은, 상기 증강 현실 시스템에서 수신된 사용자 입력에 응답하여, 상기 증강 현실 시스템의 객체 선택 디바이스에서, 상기 제1 증강 현실 환경 내의 복수의 객체들 중 상기 객체를 선택하는 것을 포함하고,

상기 복수의 객체들은 적어도 하나의 실제 객체 및 적어도 하나의 가상 객체를 포함하는,

증강 현실 시스템.

### 청구항 14

제13항에 있어서,

상기 오디오 데이터를 식별하는 것은, 별개의 사용자 입력에 응답하여, 상기 증강 현실 시스템의 상기 객체 선택 디바이스에서 선택되었던 제1 객체를 선택 해제하는 것을 더 포함하고,

상기 객체 선택 디바이스는, 상기 사용자에게 디스플레이되는 커서를 제어하는 물리적 제어기, 오디오 데이터 캡처 디바이스, 하나 이상의 음성 커맨드들을 해석하는 음성 해석 모듈, 제스처들을 추적하는 제1 시각 데이터 검출 디바이스, 상기 사용자의 눈의 위치 또는 배향을 추적하는 제2 시각 데이터 검출 디바이스, 또는 이들의 조합 중 적어도 하나를 포함하는,

증강 현실 시스템.

### 청구항 15

제11항에 있어서,

상기 증강 현실 시스템은 사용자 입력 모듈 및 위치 또는 배향 추적 모듈을 더 포함하고,

상기 증강 현실 시스템은,

상기 제1 증강 현실 환경 내의 다수의 객체들로부터 상기 객체를 선택하는 동작;

상기 증강 현실 시스템을 이용하여 상기 객체의 위치 또는 배향을 추적하는 동작;

상기 객체의 위치 또는 배향에 적어도 부분적으로 기초하여 상기 객체로부터 나오는 소리를 우선적으로 감지하는 동작

을 포함하는 동작들의 세트를 수행하도록 추가로 구성되고,

상기 오디오 데이터는, 상기 제1 증강 현실 환경에서 제1 방향, 제1 객체 위치, 또는 제1 객체 자세에서 다른 객체보다 상기 객체에 대해 우선적으로 레코딩되는,

증강 현실 시스템.

### 청구항 16

제15항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 동작은,

상기 증강 현실 시스템의 위치 및 포즈 추적 디바이스에서, 상기 제1 증강 현실 환경 내의 상기 객체에 대한 상기 제1 사용자 또는 상기 제1 사용자의 부분의 위치 및 포즈 또는 배향을 지속적으로 그리고 우선적으로 검출하는 동작;

상기 객체에 대한 상기 제1 사용자 또는 상기 제1 사용자의 일부의 위치 및 포즈 또는 배향을 지속적으로 그리고 우선적으로 검출한 결과에 적어도 부분적으로 기초하여, 상기 제1 사용자에 관한 상기 제1 메타데이터를 업데이트된 제1 메타데이터로 업데이트하는 동작; 및

상기 업데이트된 제1 메타데이터에 적어도 부분적으로 기초하여 상기 제1 로컬화 메타데이터를 업데이트하는 동작

을 더 포함하는,

증강 현실 시스템.

#### 청구항 17

제16항에 있어서,

상기 공간화된 오디오는, 상기 제2 증강 현실 환경 내의 제1 실제 객체 위치에 있는 제1 실제 객체에 의해 반사되거나, 차폐되거나 또는 가려진 것처럼 상기 제2 증강 현실 환경 내의 상기 제1 사용자 또는 제2 사용자에 의해 지각되는,

증강 현실 시스템.

#### 청구항 18

제11항에 있어서,

상기 제1 사용자의 레코딩된 증강 현실 경험을 재생하는 동작은, 상기 공간화된 오디오가 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에 의해, 상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에 대해 상대적인 제1 가상 객체 위치에 있는 제1 가상 객체로부터 나오는 것처럼 지각되도록, 상기 공간화된 오디오를 상기 제2 증강 현실 환경 내의 상기 제1 가상 객체 위치에 있는 상기 제1 가상 객체에 대해 렌더링하는 동작을 더 포함하고,

상기 오디오 데이터는 상기 제1 증강 현실 환경 내의 상기 객체로부터 나오는 것처럼 캡처되었고, 상기 객체는 상기 증강 현실 시스템에 의해 상기 제2 증강 현실 환경 내의 상기 제1 가상 객체로서 렌더링되는,

증강 현실 시스템.

#### 청구항 19

제11항에 있어서,

상기 제1 증강 현실 환경에서 제1 사용자의 증강 현실 경험을 레코딩하는 동작은,

적어도 오디오 프로세싱 모듈을 이용하여 상기 제1 환경에서 상기 오디오 데이터를 캡처하는 동작 — 상기 오디오 프로세싱 모듈은 마이크로폰, 지연 엘리먼트, 및 이득 엘리먼트를 포함하고, 상기 증강 현실 시스템에 저장된 극 패턴(polar pattern)에 적어도 부분적으로 기초하여 동작함 —;

상기 증강 현실 시스템의 이미지 캡처 디바이스를 이용하여 이미지 또는 비디오 콘텐츠 데이터를 캡처하는 동작;

상기 증강 현실 시스템의 추적 디바이스 또는 위치결정 디바이스를 이용하여 상기 이미지 또는 비디오 콘텐츠 데이터와 연관된 이미지 또는 비디오 데이터를 캡처하는 동작; 및

상기 오디오 데이터에 관련된 오디오 메타데이터를 결정하는 동작 — 상기 오디오 메타데이터를 결정하는 동작은 적어도 상기 오디오 데이터를 상기 이미지 또는 비디오 콘텐츠 데이터 또는 상기 이미지 또는 비디오



메타데이터와 연관시킴으로써 수행됨 -

을 더 포함하는,

증강 현실 시스템.

## 청구항 20

제11항에 있어서,

상기 제1 메타데이터를 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 동작은,

상기 제1 환경에서 상기 제1 사용자의 제1 사용자 위치 또는 제1 포즈에 대한 상기 객체의 위치 또는 배향에 관련된 상대 위치 또는 배향 데이터를 식별하는 동작;

상기 제1 사용자 위치에 있는 상기 제1 사용자에게 의해 지각되는 상기 오디오 데이터를 캡처하는 동작 - 상기 오디오 데이터를 캡처하는 동작은 적어도, 상기 제1 환경에서 상기 증강 현실 시스템의 마이크로폰 엘리먼트, 지연 엘리먼트, 이득 엘리먼트, 및 극 방향성 패턴을 이용함으로써 수행됨 -;

상기 상대 위치 또는 배향 데이터를 상기 오디오 데이터와 연관시키는 동작;

상기 제2 증강 현실 환경 내의 상기 제2 사용자 위치에서의 위치 또는 배향 특성을 식별하는 동작; 및

상기 제1 메타데이터를 상기 제2 사용자 위치에 있는 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 동작 - 상기 제1 메타데이터를 상기 제1 사용자 또는 제2 사용자에게 대해 로컬화하는 동작은 적어도, 상기 제2 사용자 위치에서의 상기 위치 또는 배향 특성에 적어도 부분적으로 기초하여 상기 상대 위치 또는 배향 데이터를 상기 제1 사용자 또는 제2 사용자의 머리에 적용함으로써 수행됨 -

을 포함하는,

증강 현실 시스템.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 일반적으로 가상 현실 및 증강 현실 시스템들에 관한 것이다.

### 배경 기술

[0002] 현대 컴퓨팅 및 디스플레이 기술들은 소위 "가상 현실" 또는 "증강 현실" 경험들을 위한 혼합 현실 시스템들의 개발을 용이하게 했으며, 여기서 디지털적으로 재생된 이미지들 또는 이미지들의 부분들은, 그들이 실체인 것으로 보이거나, 실제로서 지각될 수 있는 방식으로 사용자에게 제시된다. 가상현실(또는 "VR") 시나리오의 통상적으로 실제 세계 시각 입력에 대한 투과성(transparency) 없는 디지털 또는 가상 이미지 정보의 프리젠테이션과 관련된다. 증강 현실(또는 "AR") 시나리오는 통상적으로 사용자 주위의 실제 세계의 시각화에 대한 증강(즉, 다른 실제 세계 시각적 입력에 대한 투과성)으로서 디지털 또는 가상 이미지 정보의 프리젠테이션과 관련된다. 따라서, AR 시나리오들은 다른 실제 세계 시각적 입력에 대한 투과성을 동반한 디지털 또는 가상 이미지 정보의 프리젠테이션과 관련된다.

[0003] 예컨대, 도 1을 참조하면, 증강 현실 장면(scene)(4)이 도시되며, 여기서 AR 기술의 사용자는 배경에 있는 사람들, 나무들, 빌딩들, 및 콘크리트 플랫폼(8)을 피쳐링(featuring)하는 실제 세계 공원-형 세팅(6)을 본다. 이들 아이템들에 더하여, AR 기술의 최종 사용자는 또한, 자신이 실제 세계 플랫폼(8) 상에 서 있는 로봇 동상(10), 및 호박벌의 의인화인 것으로 보여지는 날고 있는 만화-형 아바타 캐릭터(12)를 보는 것을 지각하더라도, 이들 엘리먼트들(10, 12)은 실제 세계에 존재하지 않는다. 밝혀진 바와 같이, 인간 시각적 지각 시스템은 매우 복잡하고, 다른 가상 또는 실제 세계 이미지리 엘리먼트들 사이에서 가상 이미지 엘리먼트들의 편안하고, 자연스럽고, 풍부한 프리젠테이션을 가능하게 하는 VR 또는 AR 기술을 생성하는 것은 난제이다.

[0004] VR 및 AR 시스템들은 통상적으로 사용자의 머리에 적어도 느슨하게 커플링된 머리-착용 디스플레이들(또는 헬멧-장착 디스플레이들 또는 스마트 안경)을 사용하고, 이에 따라 최종 사용자의 머리가 움직일 때 움직

인다. 최종 사용자의 머리 움직임들이 디스플레이 시스템에 의해 검출되는 경우, 디스플레이되는 데이터는 머리 포즈(즉, 사용자의 머리의 배향(orientation) 및/또는 위치)의 변화를 고려하여 업데이트될 수 있다. AR(즉, 가상 및 실제 객체들의 동시성 뷰잉)을 가능하게 하는 머리-착용 디스플레이들은 몇개의 상이한 유형의 구성들을 가질 수 있다. "비디오 시-스루(video see-through)" 디스플레이로서 종종 지칭되는 하나의 이러한 구성에서, 카메라는 실제 장면의 엘리먼트들을 캡처하고, 컴퓨팅 시스템은 캡처된 실제 장면 상에 가상 엘리먼트들을 중첩하며, 불투명 디스플레이는 합성 이미지를 눈들에 제시한다. 다른 구성은 "광학 시-스루(optical see-through)" 디스플레이로서 종종 지칭되며, 여기서 최종 사용자는 환경의 실제 객체들로부터의 광을 직접 보기 위해 디스플레이 시스템의 투명(또는 반투명) 엘리먼트들을 통해 볼 수 있다. "결합기(combiner)"로서 종종 지칭되는 투명 엘리먼트는 실세계의 최종 사용자의 뷰 위에 디스플레이로부터의 광을 중첩한다.

[0005] 종종, VR/AR 시스템의 사용자는 (예컨대, 게임을 플레이하거나, 전화 회의를 하거나 또는 영화를 시청할 때) 온-라인으로 후속 게시를 위해 VR/AR 시스템 상에서의 자신의 경험을 레코딩 및 저장함으로써 그 경험을 다른 사람들과 공유하기를 원할 수 있다. 그러나, 통상적으로 시끄러운 환경으로 인해 레코딩에 노이즈 및 다른 원치 않거나 예기치 않은 사운드들이 존재하거나, 그렇지 않으면 경험에 산만함을 야기하는 너무 많은 사운드 소스들이 존재할 수 있다. 이러한 원치 않는/예기치 않은 사운드들은 실제 객체들로부터, 예컨대 VR/AR 시스템 부근에서 노는 어린이로부터 또는 가상 객체들로부터, 예컨대 VR/AR 시스템의 맥락에서 플레이되는 가상 텔레비전으로부터 유래할 수 있다.

[0006] 따라서, 사용자가 관심이 있는 가상 또는 실제 객체들로부터의 사운드들만을 레코딩하기 위한 간단하고 효율적인 수단을 제공할 필요성이 남아있다.

### 발명의 내용

[0007] 본 발명들의 제1 양상에 따라, 최종 사용자에게 의한 사용을 위한 가상 이미지 생성 시스템은, 메모리, 디스플레이 서브시스템, 및 최종 사용자로부터 입력을 수신하고, 최종 사용자 입력에 대한 응답으로 적어도 하나의 객체(예컨대, 실제 객체 및/또는 가상 객체)를 지속적으로 선택하도록 구성된 객체 선택 디바이스를 포함한다. 일 실시예에서, 디스플레이 서브시스템은 시야를 가지며, 객체 선택 디바이스는 시야에서 객체(들)를 지속적으로 선택하도록 구성된다. 이 경우에, 객체 선택 디바이스는 디스플레이 서브시스템의 시야에서 3-차원 커서를 이동시키고 최종 사용자 입력을 수신한 것에 대한 응답으로 객체(들)를 선택하도록 구성될 수 있다. 다른 실시예에서, 최종 사용자 입력은 하나 이상의 음성 커맨드들을 포함하고, 객체 선택 디바이스는 음성 커맨드(들)를 감지하도록 구성된 하나 이상의 마이크로폰들을 포함한다. 또 다른 실시예에서, 최종 사용자 입력은 하나 이상의 손 제스처들을 포함하고, 이 경우에, 객체 선택 디바이스는 손 제스처(들)를 감지하도록 구성된 하나 이상의 카메라들을 포함할 수 있다.

[0008] 복수의 객체들이 선택되는 경우에, 객체 선택 디바이스는 최종 사용자 입력에 대한 응답으로 객체들을 개별적으로 선택하고 그리고/또는 포괄적으로(globally) 선택하도록 구성될 수 있다. 포괄적으로 선택되는 경우, 객체 선택 디바이스는 최종 사용자 입력에 대한 응답으로 시야의 각도 범위 내의 모든 객체들을 포괄적으로 선택하도록 구성될 수 있다(시야의 각도 범위는 시야의 전체 각도 범위 보다 작을 수 있거나 또는 시야의 전체 각도 범위일 수 있음). 일 실시예에서, 객체 선택 디바이스는 추가로, 최종 사용자로부터 다른 입력을 수신하고 다른 최종 사용자 입력에 대한 응답으로 이전에 선택된 객체(들)를 지속적으로 선택해제(deselecting)하도록 구성된다.

[0009] 가상 이미지 생성 시스템은, 적어도 하나의 선택된 객체로부터 유래한 비디오 데이터를 생성하고, 비디오 데이터로부터 3-차원 장면으로 복수의 이미지 프레임들을 렌더링하고, 이미지 프레임들을 디스플레이 서브시스템에 전달하도록 구성된 제어 서브시스템을 더 포함한다. 일 실시예에서, 디스플레이 서브시스템은 최종 사용자의 눈들의 전방에 포지셔닝되도록 구성된다. 다른 실시예에서, 디스플레이 서브시스템은 프로젝션 서브시스템 및 부분적으로 투명한 디스플레이 표면을 포함한다. 이 경우에, 프로젝션 서브시스템은 이미지 프레임들을 부분적으로 투명한 디스플레이 표면 상에 프로젝팅하도록 구성될 수 있고, 부분적으로 투명한 디스플레이 표면은 최종 사용자의 눈들과 주변 환경 사이의 시야에 포지셔닝되도록 구성될 수 있다. 가상 이미지 생성 시스템은 최종 사용자에게 의해 착용되고 디스플레이 서브시스템의 적어도 일부를 지지하도록 구성되는 프레임 구조를 더 포함할 수 있다.

[0010] 제어 서브시스템은 추가로, 선택된 객체(들)로부터 유래하는 오디오 데이터를 생성하고, 오디오 데이터를 메모리 내에 저장하도록 구성된다. 가상 이미지 생성 시스템은 복수의 스피커들을 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 생성된 오디오 데이터를 스피커들에 전달하도록 구성될 수 있다. 선택적인

실시예에서, 제어 서브시스템은 추가로, 오디오 데이터와 동기하여 비디오 데이터를 메모리에 저장하도록 구성된다. 또 다른 실시예에서, 가상 이미지 생성 시스템은 디스플레이 서브시스템의 시야에 대한 선택된 객체(들)의 위치를 추적하도록 구성된 적어도 하나의 센서를 더 포함한다. 이 경우에, 제어 서브시스템은, 선택된 객체(들)의 추적된 위치가 디스플레이 서브시스템의 시야 밖으로 이동할 때 오디오 데이터를 메모리에 저장하는 것을 중단하도록 구성될 수 있거나 또는 대안적으로, 선택된 객체(들)의 추적된 위치가 디스플레이 서브시스템의 시야 밖으로 이동할 때 오디오 데이터를 메모리에 저장하는 것을 지속하도록 구성된다.

[0011] 선택된 객체(들)가 실제 객체를 포함하는 경우, 가상 이미지 생성 시스템은 오디오 출력을 생성하도록 구성된 마이크로폰 조립체를 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 선택된 실제 객체로부터 유래하는 사운드들을 우선적으로 감지하기 위해 오디오 출력의 방향을 수정하도록 구성될 수 있다. 오디오 데이터는 수정된 오디오 출력으로부터 도출될 수 있다. 가상 이미지 생성 시스템은 선택된 실제 객체로부터 유래하는 비디오 데이터를 캡처하도록 구성된 하나 이상의 카메라들을 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 오디오 데이터와 동기하여 비디오 데이터를 메모리에 저장하도록 구성될 수 있다. 제어 서브시스템은 캡처된 비디오 데이터를 선택된 실제 객체에 대한 가상 콘텐츠 데이터로 변환하고, 가상 콘텐츠를 메모리에 저장하도록 구성될 수 있다.

[0012] 선택된 객체(들)가 가상 객체를 포함하는 경우, 가상 이미지 생성 시스템은 복수의 가상 객체들에 대한 사운드들에 대응하는 콘텐츠 데이터를 저장하도록 구성된 데이터베이스를 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 데이터베이스로부터 선택된 가상 객체에 대응하는 콘텐츠 데이터를 획득하도록 구성될 수 있고, 메모리에 저장된 오디오 데이터는 획득된 콘텐츠 데이터를 포함한다. 제어 서브시스템은 추가로, 선택된 가상 객체에 대응하는 메타 데이터(예컨대, 선택된 가상 객체에 대한 포지션, 방향 및 볼륨 데이터)를 생성하도록 구성될 수 있으며, 이 경우에, 메모리에 저장된 오디오 데이터는 획득된 콘텐츠 데이터 및 생성된 메타 데이터를 포함할 수 있다. 일 실시예에서, 가상 이미지 생성 시스템은 최종 사용자의 머리 포즈를 추적하도록 구성된 하나 이상의 센서들을 더 포함하며, 이 경우에, 데이터베이스는 복수의 가상 객체들에 대한 절대 메타 데이터를 저장하도록 구성될 수 있고, 제어 서브시스템은 추가로, 선택된 가상 객체에 대응하는 절대 메타 데이터를 획득함으로써 메타 데이터를 생성하고 최종 사용자의 추적된 머리 포즈에 기초하여 최종 사용자에게 대해 절대 메타 데이터를 로컬화하도록 구성될 수 있다.

[0013] 가상 이미지 생성 시스템은 적어도 하나의 스피커를 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 메모리로부터 저장된 오디오 데이터를 리트리브(retrieve)하고, 리트리브된 오디오 데이터로부터 오디오를 도출하고, 오디오를 스피커(들)에 전달하도록 구성될 수 있다. 메모리에 저장된 오디오 데이터는 콘텐츠 데이터 및 메타 데이터를 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 메모리로부터 저장된 콘텐츠 데이터 및 메타 데이터를 리트리브하고, 리트리브된 콘텐츠 데이터 및 메타 데이터에 기초하여 공간화된 오디오를 렌더링하고, 렌더링된 공간화된 오디오를 스피커(들)에 전달하도록 구성될 수 있다.

[0014] 본 발명들의 제2 양상에 따라, 최종 사용자에게 의해 가상 이미지 생성 시스템을 동작시키는 방법이 제공된다. 방법은 적어도 하나의 객체(예컨대, 실제 객체 및/또는 가상 객체)를 지속적으로 선택하는 단계를 포함한다. 일 방법에서, 객체(들)를 선택하는 단계는 최종 사용자의 시야에서 3-차원 커서를 이동시키고 3-차원 커서로 객체(들)를 선택하는 단계를 포함한다. 다른 방법에서, 객체(들)를 선택하는 단계는 하나 이상의 음성 커맨드를 발행하는 단계를 포함한다. 또 다른 방법에서, 적어도 하나의 객체를 선택하는 단계는 하나 이상의 손제스처들을 만드는 단계를 포함한다. 복수의 객체들이 선택되는 경우, 복수의 객체들을 선택하는 단계는 객체들을 개별적으로 선택하는 단계 및/또는 객체들을 포괄적으로 선택하는 단계를 포함할 수 있다. 포괄적으로 선택되는 경우, 객체들은 최종 사용자의 시야의 각도 범위(이는 시야의 전체 각도 범위보다 작을 수 있거나 또는 시야의 전체 각도 범위일 수 있음)를 정의함으로써 그리고 최종 사용자의 시야의 정의된 각도 범위 내의 모든 객체들을 선택함으로써 선택될 수 있다. 선택적인 방법은 이전에 선택된 객체(들)를 지속적으로 선택해제하는 단계를 더 포함할 수 있다.

[0015] 방법은, 선택된 객체(들)로부터 유래하는 비디오 데이터를 생성하는 단계, 생성된 비디오 데이터로부터 3-차원 장면으로 복수의 이미지 프레임들을 렌더링하는 단계, 및 최종 사용자에게 이미지 프레임들을 디스플레이하는 단계, 적어도 하나의 선택된 객체로부터 유래하는 오디오 데이터를 생성하는 단계, 및 적어도 하나의 선택된 객체로부터 유래하는 오디오 데이터를 메모리 내에 저장하는 단계를 더 포함한다. 하나의 방법은 선택된 객체(들)로부터 유래하는 오디오 데이터를 최종 사용자에게 의한 지각을 위한 사운드로 변환하는 단계를 더 포함할 수 있다. 방법은 선택적으로, 오디오 데이터와 동기하여 비디오 데이터를 메모리에 저장하는 단계를 포함할 수 있다. 또 다른 방법은 최종 사용자의 시야에 대한 선택된 객체(들)의 위치를 추적하는 단계를 더 포함할 수

있다. 이 경우에, 방법은 선택된 객체(들)의 추적된 위치가 최종 사용자의 시야 밖으로 이동할 때 오디오 데이터를 메모리에 저장하는 것을 중단하는 단계 또는 대안적으로, 선택된 객체(들)의 추적된 위치가 최종 사용자의 시야 밖으로 이동할 때 오디오 데이터를 메모리에 저장하는 것을 지속하는 단계를 더 포함할 수 있다.

[0016] 선택된 객체(들)가 실제 객체를 포함하는 경우, 방법은 다른 실제 객체들로부터 유래하는 사운드들에 비해 선택된 실제 객체로부터 유래하는 사운드들을 우선적으로 감지하는 단계를 더 포함할 수 있으며, 이 경우에, 오디오 데이터는 우선적으로 감지된 사운드들로부터 도출될 수 있다. 방법은, 선택된 실제 객체로부터 유래하는 비디오 데이터를 캡처하는 단계, 및 오디오 데이터와 동기하여 비디오 데이터를 메모리에 저장하는 단계를 더 포함할 수 있다. 캡처된 비디오 데이터는 메모리에 저장하기 위해 가상 콘텐츠 데이터로 변환될 수 있다.

[0017] 선택된 객체(들)가 가상 객체를 포함하는 경우, 방법은, 복수의 가상 객체들에 대한 사운드들에 대응하는 콘텐츠 데이터를 저장하는 단계, 및 선택된 가상 객체에 대응하는 콘텐츠 데이터를 획득하는 단계를 더 포함할 수 있으며, 이 경우에, 메모리에 저장된 오디오 데이터는 획득된 콘텐츠 데이터를 포함할 수 있다. 방법은 선택된 가상 객체에 대응하는 메타 데이터(예컨대, 선택된 가상 객체에 대한 포지션, 배향 및 볼륨 데이터)를 생성하는 단계를 더 포함할 수 있으며, 이 경우에, 메모리에 저장된 오디오 데이터는 획득된 콘텐츠 데이터 및 생성된 메타 데이터를 포함할 수 있다. 방법은, 최종 사용자의 머리 포즈를 추적하는 단계, 및 복수의 가상 객체들에 대한 절대 메타 데이터를 저장하는 단계를 더 포함할 수 있다. 이 경우에, 메타 데이터를 생성하는 단계는 선택된 가상 객체에 대응하는 절대 메타 데이터를 리트리브하고 최종 사용자의 추적된 머리 포즈에 기초하여 최종 사용자에게 대해 절대 메타 데이터를 코컬화하는 단계를 포함할 수 있다.

[0018] 이 방법은 저장된 오디오 데이터를 리트리브하고, 리트리브된 오디오 데이터로부터 오디오를 도출하고, 오디오를 최종 사용자에게 의한 지각을 위한 사운드로 변환하는 단계를 더 포함할 수 있다. 저장된 오디오 데이터는 콘텐츠 데이터 및 메타 데이터를 포함할 수 있고, 이 경우에, 방법은, 메모리로부터 저장된 콘텐츠 데이터 및 메타 데이터를 리트리브하는 단계, 리트리브된 콘텐츠 데이터 및 메타 데이터에 기초하여 공간화된 오디오를 렌더링하는 단계, 및 공간화된 오디오를 최종 사용자에게 의한 지각을 위한 사운드로 변환하는 단계를 더 포함할 수 있다.

[0019] 본 발명들의 제3 양상에 따라, 플레이백 사용자에게 의한 사용을 위한 가상 이미지 생성 시스템이 제공된다. 가상 이미지 생성 시스템은, 오리지널 공간 환경에서 적어도 하나의 객체(예컨대, 실제 객체 및/또는 가상 객체)로부터 유래하는 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 저장하도록 구성된 메모리, 복수의 스피커들 및 디스플레이 서브시스템을 포함한다. 일 실시예에서, 디스플레이 서브시스템은 최종 사용자의 눈들의 전망에 포지셔닝되도록 구성된다. 다른 실시예에서, 디스플레이 서브시스템은 프로젝션 서브시스템 및 부분적으로 투명한 디스플레이 표면을 포함한다. 이 경우, 프로젝션 서브시스템은 이미지 프레임들을 부분적으로 투명한 디스플레이 표면 상에 프로젝션하도록 구성될 수 있고, 부분적으로 투명한 디스플레이 표면은 최종 사용자의 눈들과 주변 환경 사이의 시야에 포지셔닝되도록 구성될 수 있다. 가상 이미지 생성 시스템은 최종 사용자에게 의해 착용되고 디스플레이 서브시스템의 적어도 일부를 지지하도록 구성되는 프레임 구조를 더 포함할 수 있다.

[0020] 가상 이미지 생성 시스템은 메모리로부터 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 리트리브하고, 오리지널 공간 환경과 상이한 새로운 공간 환경에서 리트리브된 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터로부터의 오디오 및 비디오를 각각 렌더링하고, 렌더링된 오디오를 스피커들에 그리고 생성된 비디오 데이터를 디스플레이 서브시스템에 동기식으로 전달하도록 구성된 제어 서브시스템을 더 포함한다.

[0021] 일 실시예에서, 제어 서브시스템은 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 메모리에 저장하도록 구성된다. 가상 이미지 생성 시스템은, 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 메모리에 저장하기 전에, 최종 사용자로부터 입력을 수신하고, 최종 사용자 입력에 대한 응답으로 오리지널 공간 환경에서 객체(들)를 지속적으로 선택하도록 구성된 객체 선택 디바이스를 더 포함할 수 있다.

[0022] 객체(들)가 실제 객체를 포함하는 경우, 가상 이미지 생성 시스템은 오리지널 공간 환경 내의 실제 객체로부터 오디오 콘텐츠 데이터를 캡처하도록 구성된 마이크로폰 조립체를 더 포함할 수 있다. 마이크로폰 조립체는 오디오 출력을 생성하도록 구성될 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 선택된 실제 객체로부터 유래하는 사운드들을 우선적으로 감지하기 위해 오디오 출력의 방향을 수정하도록 구성될 수 있다. 오디오 콘텐츠 데이터는 수정된 오디오 출력으로부터 도출될 수 있다. 가상 이미지 생성 시스템은 오리지널 공간 환경 내의 선택된 실제 객체로부터 비디오 데이터를 캡처하도록 구성된 하나 이상의 카메라들을 더 포함할 수



있다. 선택적인 실시예에서, 제어 서브시스템은 캡처된 비디오 데이터를 선택된 실제 객체에 대한 가상 콘텐츠 데이터로 변환하고, 비디오 콘텐츠 데이터로서 가상 콘텐츠 데이터를 메모리에 저장하도록 구성될 수 있다.

[0023] [0023] 객체(들)가 가상 객체를 포함하는 경우, 가상 이미지 생성 시스템은 복수의 가상 객체들에 대한 사운드들에 대응하는 콘텐츠 데이터를 저장하도록 구성된 데이터베이스를 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 데이터베이스로부터 가상 객체에 대응하는 콘텐츠 데이터를 획득하도록 구성될 수 있고, 메모리에 저장된 오디오 데이터는 획득된 콘텐츠 데이터를 포함할 수 있다.

[0024] [0024] 일 실시예에서, 제어 서브시스템은 새로운 공간 환경에서 적어도 하나의 객체에 대응하는 절대 메타 데이터를 획득하고, 새로운 공간 환경에서 리트리브된 오디오 콘텐츠 데이터 및 절대 메타 데이터로부터 오디오를 렌더링하도록 구성된다. 새로운 공간 환경에서 객체(들)에 대응하는 절대 메타 데이터를 획득하는 것은 새로운 공간 환경에서 객체(들)를 포지셔닝하는 것을 포함할 수 있다. 이 경우에, 가상 이미지 생성 시스템은 플레이어백 사용자로부터 입력을 수신하도록 구성된 사용자 입력 디바이스를 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 플레이어백 사용자로부터의 입력에 대한 응답으로 새로운 공간 환경에 객체(들)를 포지셔닝하도록 구성될 수 있다. 가상 이미지 생성 시스템은 플레이어백 사용자의 머리 포즈를 추적하도록 구성된 하나 이상의 센서들을 더 포함할 수 있으며, 이 경우에, 제어 서브시스템은 추가로, 렌더링된 오디오가 공간화되도록 플레이어백 사용자의 추적된 머리 포즈에 기초하여 플레이어백 사용자에게 절대 메타 데이터를 로컬화하도록 구성될 수 있다.

[0025] [0025] 본 발명들의 제4 양상에 따라, 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터로서 오리지널 공간 환경에서 이전에 레코딩된 적어도 하나의 객체(예컨대, 실제 객체 및/또는 가상 객체)의 오디오 및 비디오를 플레이어백하기 위해 플레이어백 사용자에게 의해 가상 이미지 생성 시스템을 동작시키는 방법이 제공된다. 방법은 메모리로부터 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 리트리브하는 단계를 포함한다. 하나의 방법은 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 메모리에 저장하는 단계를 더 포함한다. 이 경우에, 방법은 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터를 메모리에 저장하기 전에 오리지널 공간 환경에서 객체(들)를 지속적으로 선택하는 단계를 더 포함할 수 있다.

[0026] [0026] 객체가 실제 객체를 포함하는 경우, 방법은 실제 객체로부터 오디오 콘텐츠 데이터를 캡처하는 단계를 더 포함할 수 있다. 이 경우에, 방법은 다른 실제 객체들로부터 유래하는 사운드들에 비해 선택된 실제 객체로부터 유래하는 사운드들을 우선적으로 감지하는 단계를 더 포함할 수 있다. 오디오 콘텐츠 데이터는 우선적으로 감지된 사운드들로부터 도출된다. 방법은, 선택된 실제 객체로부터 비디오 데이터를 캡처하는 단계, 및 캡처된 비디오 데이터를 가상 콘텐츠 데이터로 변환하는 단계를 더 포함할 수 있다. 객체가 가상 객체를 포함하는 경우, 방법은, 복수의 가상 객체들에 대한 사운드들에 대응하는 콘텐츠 데이터를 저장하는 단계, 및 데이터베이스로부터 가상 객체에 대응하는 콘텐츠 데이터를 획득하는 단계를 더 포함할 수 있다. 메모리에 저장된 오디오 콘텐츠 데이터는 획득된 콘텐츠 데이터를 포함할 수 있다.

[0027] [0027] 방법은, 오리지널 공간 환경과 상이한 새로운 공간 환경에서 리트리브된 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터로부터 오디오 및 비디오를 각각 렌더링하는 단계, 오디오 및 비디오를 사운드 및 이미지 프레임들로 각각 변환하는 단계, 및 사운드 및 이미지 프레임들을 플레이어백 사용자에게 동기식으로 전달하는 단계를 더 포함한다. 하나의 방법은 새로운 공간 환경에서 객체(들)에 대응하는 절대 메타 데이터를 획득하는 단계를 더 포함하며, 이 경우에, 오디오는 새로운 공간 환경에서 리트리브된 오디오 콘텐츠 데이터 및 절대 메타 데이터로부터 렌더링된다. 방법은, 플레이어백 사용자의 머리 포즈를 추적하는 단계, 및 플레이어백 사용자의 추적된 머리 포즈에 기초하여 플레이어백 사용자에게 절대 메타 데이터를 로컬화하는 단계를 더 포함할 수 있으며, 이 경우에, 오디오는 새로운 공간 환경에서 리트리브된 오디오 콘텐츠 데이터 및 로컬화된 메타 데이터로부터 렌더링되어서, 렌더링된 오디오가 공간화되게 할 수 있다. 새로운 공간 환경에서 객체(들)에 대응하는 절대 메타 데이터를 획득하는 것은 예컨대, 플레이어백 사용자로부터의 입력에 대한 응답으로 새로운 공간 환경에서 객체(들)를 포지셔닝하는 것을 포함할 수 있다.

[0028] [0028] 본 발명의 부가적인 그리고 다른 목적들, 특징들 및 이점들은 상세한 설명, 도면들 및 청구항들에서 설명된다.

### 도면의 간단한 설명

[0029] [0029] 도면들은 본 발명의 바람직한 실시예들의 설계 및 유용성을 예시하며, 도면들에서 유사한 엘리먼트들은 공통 참조 번호들에 의해 지칭된다. 본 발명들의 위에서 언급된 그리고 다른 이점들 및 목적들이 어떻게 달성

되는지를 더 잘 인지하기 위해, 위에서 간략하게 설명한 본 발명들의 보다 구체적 설명이 첨부 도면들에서 예시되는 본 발명의 특정 실시예들을 참조하여 제공될 것이다. 이들 도면들이 단지 본 발명의 통상적인 실시예들을 도시할 뿐이며, 이에 따라 본 발명의 범위를 제한하는 것으로 간주되지 않는다는 것을 이해하며, 본 발명은 첨부된 도면들의 사용을 통해 부가적인 특이성 및 세부사항에 관해 설명되고 기술될 것이다.

[0030] 도 1은 증강 기술의 증강 현실 생성 디바이스에 의해 최종 사용자에게 디스플레이될 수 있는 3-차원 증강 현실 장면의 사진이다.

[0031] 도 2는 본 발명들의 일 실시예에 따라 구성된 증강 현실 시스템의 사시도이다.

[0032] 도 3은 도 2의 증강 현실 시스템의 블록도이다.

[0033] 도 4는 도 2의 증강 현실 시스템에서 사용되는 공간화된 스피커 시스템의 일 실시예의 평면도이다.

[0034] 도 5는 최종 사용자가 객체를 개별적으로 선택할 수 있도록 하기 위해 도 2의 증강 현실 시스템에 의해 사용되는 하나의 기술을 예시하는 평면도이다.

[0035] 도 6은 최종 사용자가 객체를 개별적으로 선택할 수 있도록 하기 위해 도 2의 증강 현실 시스템에 의해 사용되는 다른 기술을 예시하는 평면도이다.

[0036] 도 7은 최종 사용자가 객체를 개별적으로 선택할 수 있도록 하기 위해 도 2의 증강 현실 시스템에 의해 사용되는 또 다른 기술을 예시하는 평면도이다.

[0037] 도 8은 최종 사용자가 다수의 객체를 포괄적으로 선택할 수 있도록 하기 위해 도 2의 증강 현실 시스템에 의해 사용되는 기술을 예시하는 평면도이다.

[0038] 도 9는 최종 사용자가 다수의 객체를 포괄적으로 선택할 수 있도록 하기 위해 도 2의 증강 현실 시스템에 의해 사용되는 다른 기술을 예시하는 평면도이다.

[0039] 도 10a는 도 2의 증강 현실 시스템을 착용하는 데 사용될 수 있는 하나의 기술의 평면도이다.

[0040] 도 10b는 도 2의 증강 현실 시스템을 착용하는 데 사용될 수 있는 다른 기술의 평면도이다.

[0041] 도 10c는 도 2의 증강 현실 시스템을 착용하는 데 사용될 수 있는 또 다른 하나의 기술의 평면도이다.

[0042] 도 10d는 도 2의 증강 현실 시스템을 착용하는 데 사용될 수 있는 또 다른 기술의 평면도이다.

[0043] 도 11은 다양한 예시적인 다양한 가상 및 실제 사운드와 상호작용하는 도 2의 증강 현실 시스템을 예시하는 블록도이다.

[0044] 도 12는 도 2의 증강 현실 시스템에서 사용되는 오디오 프로세서의 일 실시예를 예시하는 블록도이다.

[0045] 도 13은 도 2의 증강 현실 시스템에 의해 선택된 가상 및 실제 객체들에 대응하는 콘텐츠 데이터 및 메타 데이터를 레코딩하는 메모리의 다이어그램이다.

[0046] 도 14는 실제 객체들로부터 사운드를 우선적으로 수신하기 위해 도 2의 증강 현실 시스템에서 사용되는 마이크로폰 조립체 및 대응하는 오디오 프로세싱 모듈들의 개략도이다.

[0047] 도 15a는 최종 사용자에게 대한 제1 배향을 갖는 2개의 객체들로부터 사운드를 우선적으로 수신하기 위해 도 2의 증강 현실 시스템의 오디오 프로세서에 의해 생성된 방향성 패턴들의 평면도이다.

[0048] 도 15b는 최종 사용자에게 대한 제2 배향을 갖는 2개의 객체들로부터 사운드를 우선적으로 수신하기 위해 도 2의 증강 현실 시스템의 오디오 프로세서에 의해 생성된 방향성 패턴들의 평면도이다.

[0049] 도 16a는 최종 사용자에게 대해 오리지널 공간 환경에 분산된 객체들의 블록도이다.

[0050] 도 16b는 최종 사용자에게 대해 새로운 공간 환경에 분산된 도 17a의 객체들의 블록도이다.

[0051] 도 17은 가상 및 실제 객체들의 오디오 및 비디오를 선택 및 레코딩하기 위해 도 2의 증강 현실 시스템을 동작시키는 하나의 방법을 예시하는 흐름도이다.

[0052] 도 18은 새로운 공간 환경에서 도 17에 레코딩된 오디오 및 비디오를 플레이백하기 위해 도 2의 증강 현실 시스템을 동작시키는 하나의 방법을 예시하는 흐름도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0030] [0053] 이어지는 설명은 증강 현실 시스템에서 사용되는 디스플레이 시스템들 및 방법들에 관한 것이다. 그러나, 본 발명은 증강 현실 시스템들의 애플리케이션들에 매우 적합하지만, 본 발명은 그의 가장 넓은 양상들에서, 그러한 것으로 제한되지 않을 수 있다는 것이 이해될 것이다. 예컨대, 본 발명은 가상 현실 시스템들에 적용될 수 있다. 따라서, 종종 본원에서 증강 현실 시스템의 관점에서 설명되지만, 교시들이 그러한 용도들의 그러한 시스템들로 제한되지는 안 된다. 증강 현실 시스템은 예컨대, 비디오 게임, 가상 사람과 실제 사람의 조합으로 이루어진 원격 회의, 또는 영화 관람의 맥락에서 동작될 수 있다.
- [0031] [0054] 본원에서 설명된 증강 현실 시스템은 최종 사용자가, 최종 사용자에 의해 지속적으로 선택된 적어도 하나의 객체(가상 또는 실제)로부터 유래하는 오디오 데이터를 레코딩할 수 있게 한다. 이러한 레코딩된 오디오 데이터는 후속적으로, 동일하거나 상이한 최종 사용자에 의해 플레이백될 수 있다. 레코딩된 오디오 데이터로부터 유래하는 사운드는 오디오 데이터가 원래 레코딩된 실제 환경에서 동일하거나 상이한 최종 사용자에게 플레이백될 수 있다. 오디오 데이터의 콘텐츠를 레코딩하는 것에 더하여, 최종 사용자의 머리 포즈뿐만 아니라 오디오 콘텐츠가 원래 레코딩된 환경을 특성화하는 메타 데이터가 그러한 오디오 데이터와 관련하여 레코딩될 수 있어서, 플레이백 동안, 오디오는 최종 사용자가 오리지널 레코딩 동안 공간화된 사운드를 청각적으로 경험했던 것과 동일한 방식으로 청각적으로 경험되는 공간화된 사운드로 재-렌더링 및 변환할 수 있게 한다. 선택적으로, 오디오는 새로운 가상 또는 실제 환경에서 동일하거나 상이한 최종 사용자에 의한 지각을 위한 공간화된 사운드로 재-렌더링 및 변환될 수 있어서, 동일하거나 상이한 최종 사용자가 새로운 환경에 적절한 청각 경험을 할 수 있게 한다. 오디오 데이터는 주변 환경의 가상 객체들 및 실제 객체들로부터 유래한 비디오 데이터와 동기하여 레코딩될 수 있다.
- [0032] [0055] 본원에서 설명된 증강 현실 시스템은, 최종 사용자의 시야 내의 실제(또는 물리적) 객체들과 혼합되는 가상 객체들의 이미지들을 제공할 뿐만 아니라 (시야 내부 또는 외부의) 실제(또는 물리적) 소스들로부터 유래하는 실제 사운드와 혼합되는, (시야 내부 또는 외부의) 가상 소스들로부터 유래하는 가상 사운드를 제공하도록 동작될 수 있다. 이를 위해, 본 발명들에 따라 구성된 증강 현실 시스템(100)의 일 실시예가 이제 도 2 및 도 3을 참조로 설명될 것이다. 증강 현실 시스템(100)은 디스플레이 스크린(104) 및 이미지들을 디스플레이 스크린(104) 상으로 프로젝팅하는 프로젝션 서브시스템(도시되지 않음)을 포함하는 디스플레이 서브시스템(102)을 포함한다.
- [0033] [0056] 예시된 실시예에서, 디스플레이 스크린(104)은 주변 환경에 있는 실제 객체들이 최종 사용자(50)에 의해 보여질 수 있게 하고 가상 객체들의 이미지들이 디스플레이될 수 있는 부분적으로 투명한 디스플레이 스크린이다. 증강 현실 시스템(100)은, 디스플레이 스크린(104)이 최종 사용자(50)의 눈들(52)의 전방에 그리고 특히, 주변 환경과 최종 사용자(50)의 눈들(52) 사이의 최종 사용자(50)의 시야에 포지셔닝되도록, 부분적으로 투명한 디스플레이 스크린(104)을 지지하는, 최종 사용자(50)에 의해 착용되는 프레임 구조(106)를 더 포함한다.
- [0034] [0057] 디스플레이 서브시스템(102)은, 2-차원 콘텐츠를 제시 가능할 뿐만 아니라, 고레벨의 이미지 품질 및 3-차원 지각으로, 물리적 현실에 대한 증강들로서 편안하게 지각될 수 있는 포토-기반 방사 패턴(photo-based radiation pattern)들을 최종 사용자(50)의 눈들(52)에 제시하도록 설계된다. 디스플레이 서브시스템(102)은 단일 코히어런트(coherent) 장면의 지각을 제공하는 고주파수의 프레임들의 시퀀스를 제시한다.
- [0035] [0058] 대안적인 실시예들에서, 증강 현실 시스템(100)은 주변 환경의 이미지들을 캡처하여 비디오 데이터로 변환하기 위해 하나 이상의 이미지들(예컨대, 카메라들)를 사용할 수 있으며, 이 비디오 데이터는 그 후 가상 객체들을 표현하는 비디오 데이터와 혼합될 수 있으며, 이 경우에, 증강 현실 시스템(100)은 불투명한 디스플레이 표면 상에서 최종 사용자(50)에게 혼합된 비디오 데이터를 나타내는 이미지들을 디스플레이할 수 있다.
- [0036] [0059] 디스플레이 서브시스템들을 설명하는 추가의 세부사항들은, 발명의 명칭이 "Display Subsystem and Method"인 미국 가특허 출원 일련 번호 제14/212,961호 및 발명의 명칭이 "Planar Waveguide Apparatus With Diffraction Element(s) and Subsystem Employing Same"인 미국 가특허 출원 일련 번호 제14/331,216호에서 제공되며, 이들은 인용에 의해 본원에 명시적으로 포함된다.
- [0037] [0060] 증강 현실 시스템(100)은, 최종 사용자(50)가 실제 객체들로부터의 사운드들을 직접 들을 수 있게 하면서, 가상 객체들로부터 최종 사용자(50)에게 사운드만을 제시하기 위한 하나 이상의 스피커(들)(108)를 더 포함한다. 대안적인 실시예들에서, 증강 현실 시스템(100)은 주변 환경으로부터 유래하는 실제 사운드를 캡처하여

오디오 데이터로 변환하기 위한 하나 이상의 마이크로폰들(도시되지 않음)을 포함할 수 있으며, 이 오디오 데이터는 가상 사운드로부터의 오디오 데이터와 혼합될 수 있으며, 이 경우에 스피커(들)(108)는 혼합된 오디오 데이터를 나타내는 사운드를 최종 사용자(50)에게 전달할 수 있다.

[0038] [0061] 어느 경우든, 스피커(들)(108)는 프레임 구조(106)에 의해 지지되어서, 스피커(들)(108) 예컨대, 이어버드(earbud)들 또는 헤드폰이 최종 사용자(50)의 외이도에 인접하게(그 안에 또는 그 주변에) 포지셔닝되게 한다. 스피커(들)(108)는 스테레오/형상화 가능한 사운드 제어를 제공할 수 있다. 스피커(들)(108)가 외이도들에 인접하게 포지셔닝되는 것으로서 설명되었지만, 외이도들에 인접하게 로케이팅되지 않은 다른 유형들의 스피커들이 최종 사용자(50)에게 사운드를 전달하는 데 사용될 수 있다. 예컨대, 스피커들은 예컨대, 골전도(bone conduction) 기술을 사용하여 외이도들로부터 거리를 두고 배치될 수 있다. 도 4에 예시된 선택적인 실시예에서, 다수의 공간화된 스피커들(108)(예컨대, 4개의 스피커들(108-1, 108-2, 108-3 및 108-4))이 최종 사용자(50)의 머리(54) 주위에 로케이팅될 수 있고 머리(54)의 좌측, 우측, 전방 및 후방으로부터의 그리고 최종 사용자(50)의 좌측 및 우측 귀들(56) 쪽을 향하는 사운드를 수신하도록 구성될 수 있다. 증강 현실 시스템에 대해 사용될 수 있는 공간화된 스피커들에 관한 추가의 세부사항들은 발명의 명칭이 "Mixed Reality System with Spatialized Audio"인 미국 가특허 출원 일련 번호 제62/369,561호에서 설명되며, 이는 인용에 의해 본원에 명시적으로 포함된다.

[0039] [0062] 중요하게는, 증강 현실 시스템(100)은 단지 이들 선택된 객체(들)로부터만의 사운드들의 레코딩을 위해 최종 사용자(50)가 하나 또는 몇몇 또는 모든 객체들(가상 또는 실제)을 선택할 수 있도록 구성된다. 이를 위해, 증강 현실 시스템(100)은 최종 사용자(50)로부터의 입력에 대한 응답으로, 하나 이상의 실제 객체들(즉, 실제 사운드들이 유래하는 실제 객체들) 및 가상 객체들(즉, 가상 사운드가 유래하는 가상 객체들)을, 그들로부터의 사운드들을 레코딩하기 위해, 선택하도록 구성된 객체 선택 디바이스(110)를 더 포함한다. 객체 선택 디바이스(110)는 최종 사용자(50)의 시야 내의 실제 객체 또는 가상 객체를 개별적으로 선택하고 그리고/또는 최종 사용자(50)의 시야 내의 실제 객체들 또는 가상 객체들의 서브세트 또는 전부를 포괄적으로 선택하도록 설계될 수 있다. 객체 선택 디바이스(110)는 또한 최종 사용자(50)로부터의 부가적인 입력에 대한 응답으로 하나 이상의 이전에 선택된 실제 객체들 또는 가상 객체들을 선택해제하도록 구성될 수 있다. 이 경우에, 객체 선택 디바이스(110)는 실제 객체들 또는 가상 객체들이 이전에 선택된 것과 동일한 방식으로 이들을 선택해제하도록 설계될 수 있다. 어느 경우든, 특정 객체가 지속적으로 선택되며, 이는 의도적으로 선택해제될 때까지 특정 객체가 선택된 상태로 유지된다는 것을 의미한다.

[0040] [0063] 일 실시예에서, 디스플레이 서브시스템(102)은 최종 사용자(50)의 시야에 3-차원 커서를 디스플레이할 수 있으며, 이 3-차원 커서는, 객체 선택 디바이스(110)로의 입력에 대한 응답으로, 증강 현실 장면에서 특정 실제 객체 또는 가상 객체를 선택할 목적으로 최종 사용자(50)의 시야에서 변위될 수 있다.

[0041] [0064] 예컨대, 도 5에 도시된 바와 같이, 4개의 가상 객체들(V1-V4) 및 2개의 실제 객체들(R1-R2)이 디스플레이 이 스크린(104)의 시야(60) 내에 로케이팅된다. 디스플레이 서브시스템(102)은, 원의 형태를 취하는 것으로 예시되는 3D 커서(62)를 시야(60)에 디스플레이할 수 있다. 3D 커서(62)는, 최종 사용자(50)에 의한 객체 선택 디바이스(110)로의 입력에 대한 응답으로, 객체들 중 하나 위로 그리고 이 경우에는 가상 객체(V3) 위로 이동될 수 있으며, 그리하여 3D 커서(62)가 그 객체와 연관된다. 연관된 객체가 그 후, 최종 사용자(50)에 의한 객체 선택 디바이스(110)로의 부가적인 입력에 대한 응답으로 선택될 수 있다. 특정 객체(이 경우에, 가상 객체(V3))가 3D 커서(62)와 연관되고 선택을 위한 준비가 되었다는 시각적 피드백을 제공하기 위해, 연관된 객체, 또는 심지어 3D 커서(62) 그 자체가 강조(예컨대, 컬러의 변화 또는 음영)될 수 있다. 일단 선택되면, 객체는 그것이 선택해제될 때까지 강조된 채로 유지될 수 있다. 물론, 가상 객체(V3) 대신에 또는 가상 객체(V3)에 부가하여, 실제 객체들을 포함해서, 증강 현실 장면(4)의 다른 객체들이, 이들 다른 객체들 중 임의의 것 위에 3D 커서(62)를 배치하고 3D 커서(62) 내의 객체를 선택함으로써 선택될 수 있다. 또한, 도 5의 3D 커서(62)가 원의 형태를 취하지만, 3D 커서(62)는, 화살표를 포함하여, 특정 객체를 가리키기 위해 최종 사용자(50)에 의해 사용될 수 있는 임의의 형상일 수 있다는 것이 인지되어야 한다. 시야(60)에서 이전에 선택된 임의의 객체들 중 임의의 것은 이전에 선택된 객체 위로 3D 커서(62)를 이동시키고 이를 선택해제함으로써 선택해제될 수 있다.

[0042] [0065] 객체 선택 디바이스(110)는 최종 사용자(50)가 특정 객체 위로 3D 커서(62)를 이동시키고 후속적으로 그 특정 객체를 선택할 수 있게 하는 임의의 디바이스의 형태를 취할 수 있다. 일 실시예에서, 객체 선택 디바이스(110)는 특정 객체 위로 3D 커서(62)를 이동시키도록 물리적으로 조작되고 특정 객체를 선택하도록 "클릭"



될 수 있는 마우스, 터치패드, 조이스틱, 방향 버튼 등과 같은 종래의 물리적 제어기의 형태를 취한다.

- [0043] [0066] 다른 실시예에서, 객체 선택 디바이스(110)는 음성 커맨드들에 대한 응답으로 특정 객체 위로 3D 커서(62)를 이동시키고 그 후 특정 객체를 선택할 수 있는 마이크로폰 및 대응하는 음성 해석 모듈을 포함할 수 있다. 예컨대, 최종 사용자(50)는 특정 객체 위로 3D 커서(62)를 점진적으로 이동시키기 위해 방향 커맨드들, 예컨대, 좌측으로 이동 또는 우측으로 이동을 발언하고, 그 후 특정 객체를 선택하기 위해 "선택"과 같은 커맨드를 발언할 수 있다.
- [0044] [0067] 또 다른 실시예에서, 객체 선택 디바이스(110)는, 특정 객체의 선택을 위해 특정 객체 위로 3D 커서(62)를 상응하게 이동시키는 최종 사용자(50)에 의한 물리적 제스처(예컨대, 손가락 움직임)를 추적할 수 있는, 프레임 구조(106)에 장착된 하나 이상의 카메라들(예컨대, 전향 카메라(들)(112)), 및 대응하는 프로세서(도시되지 않음)를 포함할 수 있다. 예컨대, 최종 사용자(50)는 손가락을 사용하여 시야(60) 내에서 특정 객체 위로 3D 커서(62)를 "드래그(drag)"하고 그 후 특정 객체를 선택하도록 3D 커서(62)를 "탭핑(tap)"할 수 있다. 다시 말해, 전향 카메라(들)(112)는 예컨대, 특정 객체의 선택을 위해 특정 객체 위에 3D 커서(62)를 상응하게 이동시키는 최종 사용자(50)의 머리(54)의 배향에 적어도 부분적으로 기초하여, 예컨대, 최종 사용자(50)의 주목의 중심(center of attention)을 검출하거나 추론하기 위해 이용될 수 있다. 예컨대, 최종 사용자(50)는 시야(60) 내에서 특정 객체 위로 3D 커서(62)를 "드래그"하기 위해 자신의 머리(50)를 움직이고, 그 후 특정 객체를 선택하도록 자신의 머리(50)를 빠르게 끄덕일 수 있다.
- [0045] [0068] 또 다른 실시예에서, 객체 선택 디바이스(110)는, 최종 사용자(50)의 눈들(52) 및 특히, 최종 사용자(50)가 포커싱하는 방향 및/또는 거리를 추적하는, 하나 이상의 카메라들(예컨대, 후향 카메라(들)(114)(도 2에 도시됨)) 및 대응하는 프로세서를 포함할 수 있으며, 이는 특정 객체의 선택을 위해 그 특정 객체 위로 3D 커서(62)를 상응하게 이동시킨다. 후향 카메라(들)(114)는 최종 사용자(50)의 눈들(52)의 각 포지션(angular position)(눈 또는 눈이 가리키는 방향), 깜박임 및 (눈 수렴을 검출함으로써) 초점 깊이를 추적할 수 있다. 예컨대, 최종 사용자(50)는 특정 객체 위로 3D 커서를 "드래그"하기 위해 시야 내에서 자신의 눈들(54)을 움직이고 그 후 특정 객체를 선택하도록 깜박일 수 있다. 이러한 눈 추적 정보는 예컨대, 최종 사용자의 눈들에 광을 프로젝팅하고 그 프로젝팅된 광 중 적어도 일부의 리턴 또는 반사를 검출함으로써 인식될 수 있다. 눈 추적 디바이스들을 논의하는 추가의 세부사항들은, 발명의 명칭이 "Display Subsystem and Method"인 미국 가특허 출원 일련 번호 제14/212,961호, 발명의 명칭이 "Methods and Subsystem for Creating Focal Planes in Virtual and Augmented Reality"인 미국 특허 출원 일련 번호 제14/726,429호, 및 발명의 명칭이 "Subsystem and Method for Augmented and Virtual Reality"인 미국 특허 출원 일련 번호 제14/205,126호에서 제공되며, 이들은 인용에 의해 본원에 명시적으로 포함된다.
- [0046] [0069] 대안적인 실시예들에서, 객체 선택 디바이스(110)는 3D 커서(62)를 이동 및 사용하여 객체를 선택하기 위해 종래의 물리적 제어기, 마이크로폰/음성 해석 모듈 및/또는 카메라를 결합할 수 있다. 예컨대, 물리적 제어기, 손가락 제스처 또는 눈 움직임은 특정 객체 위로 3D 커서(62)를 이동시키는 데 사용될 수 있고, 그 특정 객체를 선택하는 데 음성 커맨드가 사용될 수 있다.
- [0047] [0070] 최종 사용자(50)의 시야에서 객체들을 선택하기 위해 3D 커서(62)를 사용하기 보다는, 특정 객체는 그 특정 객체를 의미적으로(semantically) 식별하거나 최종 사용자(50)에게 디스플레이되는 메뉴를 통해 객체를 선택함으로써 선택될 수 있으며, 이 경우에 객체는 최종 사용자(50)의 시야에 있을 필요는 없다. 이 경우에, 객체 선택 디바이스(110)는 마이크로폰, 및 특정 객체가 의미적으로 식별되는 경우, 최종 사용자(50)에 의해 제공되는 구두 커맨드들을 번역하는 음성 해석 모듈의 형태를 취한다. 예컨대, 가상 객체(V3)가 드럼들에 대응하는 경우, 최종 사용자(50)는 "드럼들을 선택해"라고 발언할 수 있고, 이에 대한 응답으로 드럼들(V3)이 선택될 것이다. 구두 커맨드에 대응하는 객체의 선택을 용이하게 하기 위해, 시야 내의 모든 관련 객체들을 식별하는 시멘틱 정보가 바람직하게는 데이터베이스에 저장되어서, 최종 사용자(50)에 의해 구두로 표현된 객체의 설명이 데이터베이스에 저장된 객체의 설명과 매칭될 수 있다. 시멘틱 정보를 포함하는 메타 데이터는 데이터베이스의 가상 객체들과 사전에 연관될 수 있는 반면, 시야 내의 실제 객체들은 발명의 명칭이 "Method and System for Inserting Recognized Object Data into a Virtual World"인 미국 특허 출원 일련 번호 제14/704,800호(이는 명시적으로 인용에 의해 포함됨)에서 설명된 방식으로 시멘틱 정보와 사전에 매핑 및 연관될 수 있다.
- [0048] [0071] 대안적으로, 특정 객체는, 손가락 제스처를 사용하여 단순히 그 특정 객체를 가리키거나 "클릭"함으로써, 3D 커서(62)를 사용하지 않고 특정 객체가 선택될 수 있다. 이 경우에, 객체 선택 디바이스(110)는 특정 객체의 선택을 위한 손가락 제스처를 추적하는, 하나 이상의 카메라들(예컨대, 전향 카메라들

(114)) 및 대응하는 프로세서를 포함할 수 있다. 예컨대, 최종 사용자(50)는, 도 6에 도시된 바와 같이 단순히 특정 객체(이 경우 가상 객체(V3))를 가리킴으로써 그 특정 객체를 선택할 수 있다. 다른 실시예에서, 특정 객체는, 도 7에 도시된 바와 같이 적어도 2개의 손가락들(예컨대, 검지 손가락 및 엄지 손가락)을 사용하여 원 또는 부분 원을 형성함으로써 3D 커서(62)를 사용하지 않고 선택될 수 있다.

[0049] [0072] 3D 커서(62)는 한 번에 단 하나의 객체만을 선택하도록 사용되는 것으로 설명되었지만, 대안적 또는 선택적인 실시예들에서, 3D 커서(62)는 한 번에 다수의 객체들을 선택하도록 사용될 수 있다. 예컨대, 도 8에 예시된 바와 같이, 3D 커서(62)를 사용하여 객체들의 그룹 주위, 예컨대 실제 객체(R1) 및 가상 객체들(V3 및 V4) 주위에 라인(64)이 그려질 수 있으며, 그리하여 이들 객체들의 그룹이 선택된다. 3D 커서(62)는 예컨대, 객체들을 개별적으로 선택하기 위해 위에서 설명된 동일한 수단을 사용하여 제어될 수 있다. 대안적으로, 3D 커서(62)를 사용하지 않고, 예컨대 손가락 제스처를 사용함으로써 객체들의 그룹 주위에 라인이 그려질 수 있다.

[0050] [0073] 대안적인 실시예에서, 최종 사용자(50)의 시야의 미리 정의된 각도 범위 내의 객체들의 그룹이 선택될 수 있으며, 이 경우에, 객체 선택 디바이스(110)는 예컨대, 이들 객체들을 선택하기 위해 최종 사용자(50)에 의해 작동될 수 있는 단일 물리적 또는 가상 선택 버튼의 형태를 취할 수 있다. 시야의 각도 범위는 최종 사용자(50)에 의해 이전에 미리 정의될 수 있거나 증강 현실 시스템(100) 내에 미리 프로그래밍될 수 있다. 예컨대, 도 9에 도시된 바와 같이, 60도(시야 중심으로부터  $\pm 30$ 도)의 각도 범위(66)가 120 도의 시야(60)의 맥락에서 도시된다. 시야(60)의 각도 범위(64) 내의 모든 객체들(이 경우에, 가상 객체들(V1, V2 및 V3))은 선택 버튼의 작동 시에 포괄적으로 선택될 수 있는 반면, 시야(60)의 각도 범위(64) 외부의 모든 객체들(이 경우에, 실제 객체들(R1 및 R2), 및 가상 객체(V4))은 선택 버튼의 작동 시에 선택되지 않을 것이다. 일 실시예에서, 최종 사용자(50)는, 예컨대 각도 범위를 정의하는 예지들 중 하나 또는 둘 모두를, 시야(60)의 중심선을 향하거나 멀어지게 드래그(화살표들에 의해 표시됨)함으로써 각도 범위를 수정할 수 있다. 최종 사용자(50)는 예컨대, 최소 0도 내지 전체 시야(예컨대, 120도)의 각도 범위를 조정할 수 있다. 대안적으로, 시야(60)의 각도 범위(64)는 최종 사용자(50)가 그것을 조정하는 능력 없이 미리 프로그래밍될 수 있다. 예컨대, 전체 시야(60) 내의 모든 객체들이 선택 버튼의 작동에 대한 응답으로 선택될 수 있다.

[0051] [0074] 증강 현실 시스템(100)은 주변 환경의 실제 객체들로부터의 사운드들을 오디오 신호들로 변환하도록 구성된 하나 이상의 마이크로폰들을 더 포함한다. 특히, 증강 현실 시스템(100)은 객체 선택 디바이스(110)를 통해 최종 사용자(50)에 의해 선택된 하나 이상의 실제 객체들의 방향 및 거리에 대응하는 특정 방향 및/또는 특정 거리의 사운드를 우선적으로 수신하도록 구성된 마이크로폰 조립체(116)를 포함한다. 마이크로폰 조립체(116)는 도 2에 예시된 바와 같이(단 2개만이 도시됨) 프레임 구조(106)에 장착된 마이크로폰 엘리먼트들(118)(예컨대, 4개의 마이크로폰)의 어레이를 포함한다. 마이크로폰 조립체(116)에 관한 세부사항들은 아래에서 추가로 상세히 설명될 것이다. 증강 현실 시스템(100)은 최종 사용자(50)의 스피치를 오디오 신호들로 변환하도록, 예컨대 최종 사용자(50)로부터 커맨드들 또는 내레이션을 수신하도록 구성된 전용 마이크로폰(122)을 더 포함한다.

[0052] [0075] 증강 현실 시스템(100)은 알려진 좌표계 내에서 선택된 실제 객체(들)의 포지션 및 배향을 추적하여서, 이들 실제 객체(들)로부터 유래하는 사운드들은, 증강 현실 시스템들에 대한 선택된 실제 객체(들)의 포지션 또는 배향이 변하는 동안에도 마이크로폰 조립체(116)에 의해 선택되지 않은 실제 객체(들)에 비해 우선적으로 그리고 지속적으로 감지될 수 있게 한다. 알려진 좌표계에서 모든 가상 객체들의 포지션 및 위치는 통상적으로, 증강 현실 시스템(100)에 "알려져 있고"(즉, 그 내에 레코딩됨), 따라서 일반적으로 능동적으로 추적될 필요가 없다.

[0053] [0076] 예시된 실시예에서, 증강 현실 시스템(100)은 사운드의 선명도 또는 사실감(realism)에 영향을 미치도록 실제 및 물리적 3-차원(3D) 공간에서 가상 위치들 및 배향들이 알려진 가상 객체들에 대응하는 공간화된 오디오를 렌더링 및 제시하는 공간화된 오디오 시스템을 이용하여 실제 객체들의 가상 위치들로부터 사운드들이 유래하는 것처럼 최종 사용자(50)에게 나타나게 한다. 증강 현실 시스템(100)은 다양한 가상 객체들과 연관된 오디오가 그들의 가상 포지션들로부터 유래하는 것처럼 나타나게 하도록 공간화된 오디오를 보다 정확하게 렌더링하기 위해 최종 사용자(50)의 포지션을 추적한다. 또한, 증강 현실 시스템(100)은 다양한 가상 객체들과 연관된 방향성 오디오가 각각의 가상 객체들에 적합한 가상 방향으로 전파되는 것처럼(예컨대, 가상 캐릭터의 머리 뒤에서 나오는 것이 아니라 가상 캐릭터의 입으로부터 나온) 나타나도록, 공간화 오디오를 보다 정확하게 렌더링하기 위해 최종 사용자(50)의 머리 포즈를 추적한다. 또한, 증강 현실 시스템(100)은 다양한 가상 객체들과 연관된 오디오가 실제 물리적 및 가상 객체들로부터 적절히 반사되거나 또는 이들에 의해 차폐되거나 가려지

는 것처럼 나타나도록, 공간화 오디오를 렌더링하는데 있어 다른 실제 물리적 및 가상 객체들을 고려한다.

- [0054] [0077] 이를 위해, 증강 현실 시스템(100)은 가상 3-차원 장면에 대한 최종 사용자(50)의 머리(54)의 포지션 및 배향을 추적할 뿐만 아니라 최종 사용자(50)의 머리(54)에 대한 실제 객체들의 포지션 및 배향을 추적하기 위한 머리/객체 추적 서브시스템(120)을 더 포함한다. 예컨대, 머리/객체 추적 서브시스템(120)은 최종 사용자(50)의 머리 포즈 데이터(위치 및 배향)를 수집하도록 구성된 하나 이상의 센서들, 및 센서(들)(120)에 의해 수집된 머리 포즈 데이터에 기초하여 알려진 좌표계에서 최종 사용자(50)의 머리 포즈를 결정하도록 구성된 프로세서(도시되지 않음)를 포함할 수 있다. 센서(들)는 하나 이상의 이미지 캡처 디바이스(이를테면, 가시 및 적외선 카메라), 관성 측정 유닛들(가속도계들 및 자이로스코프들을 포함함), 컴퍼스들, 마이크로폰들, GPS 유닛들 또는 라디오 디바이스들 중 하나 이상을 포함할 수 있다. 예시된 실시예에서, 센서(들)는 전향 카메라(들)(112)(도 2에 도시됨)를 포함한다. 이러한 방식으로 머리에 착용될 때, 전향 카메라(들)(120)은 최종 사용자(50)가 로케이팅되는 환경에 대해 최종 사용자(50)의 머리(54)의 거리 및 각 포지션(즉, 머리가 가리키는 방향)을 나타내는 정보를 캡처하는데 특히 적합하다. 머리 배향은 임의의 방향(예컨대, 최종 사용자(50)의 기준 프레임에 대해 위/아래, 좌, 우)에서 검출될 수 있다. 아래에 추가로 상세히 설명될 바와 같이, 전향 카메라(들)(114)는 또한 증강 현실 시스템(100)의 비디오 레코딩 기능을 용이하게 하기 위해 주변 환경의 실제 객체들의 비디오 데이터를 획득하도록 구성된다. 또한, 주변 환경에서 실제 객체를 추적하기 위해 카메라들이 제공될 수 있다. 프레임 구조(106)는 카메라들이 프레임 구조(106)의 전방 및 후방에 장착될 수 있도록 설계될 수 있다. 이러한 방식으로, 카메라들의 어레이는 관련 객체들의 모든 방향들을 커버하기 위해 최종 사용자(50)의 머리(54)를 둘러쌀 수 있다.
- [0055] [0078] 증강 현실 시스템(100)은, 가상 객체들(가상 객체들의 콘텐츠 데이터뿐만 아니라 이들 가상 객체들과 연관된 절대 메타 데이터, 예컨대, 3-차원 장면에서 이들 가상 객체들의 절대 포지션 및 배향) 및 가상 객체들을 포함하는 3D 장면(가상 객체들의 콘텐츠 데이터뿐만 아니라 이들 가상 객체들과 연관된 절대 메타 데이터, 예컨대, 3D 장면에서 이들 가상 객체들의 볼륨 및 절대 포지션 및 배향뿐만 아니라 가상 소스 부근의 임의의 가상 또는 실제 객체들을 포함한 각각의 가상 객체들을 둘러싼 공간 음향, 룸 치수들, 벽/바닥 재료들 등)을 저장하도록 구성된 3-차원 데이터베이스(124)를 더 포함한다.
- [0056] [0079] 증강 현실 시스템(100)은, 시야에 나타나는 실제 객체들 및 가상 객체들로부터 유래하는 비디오 데이터를 레코딩하는 것 외에도, 객체 선택 디바이스(110)를 통해 최종 사용자(50)가 선택한 그러한 가상 객체들 및 실제 객체들로부터만 유래하는 오디오 데이터를 레코딩하는 제어 서브시스템을 더 포함한다. 증강 현실 시스템(100)은 또한, 비디오 데이터 및 오디오 데이터와 연관된 메타 데이터를 레코딩할 수 있어서, 동기화된 비디오 및 오디오가 플레이백 동안 정확하게 재-렌더링될 수 있게 한다.
- [0057] [0080] 이를 위해, 제어 서브시스템은 3-차원 데이터베이스(124)로부터 가상 객체들과 연관된 비디오 콘텐츠 및 절대 메타 데이터를 획득하고 머리/객체 추적 서브시스템(120)으로부터 (아래에서 추가로 상세히 설명될 바와 같이, 비디오에 대한 절대 메타 데이터를 최종 사용자(50)의 머리(54)에 대해 로컬화하는 데 사용될) 최종 사용자(50)의 머리 포즈 데이터를 획득하고, 그로부터 비디오를 렌더링하도록 구성된 비디오 프로세서(126)를 포함하며, 이 비디오는 그 후, 최종 사용자(50)의 시야 내의 주변 환경의 실제 객체들로부터 유래하는 이미지들과 혼합되는 이미지들로의 변환을 위해 디스플레이 서브시스템(102)으로 전달된다. 비디오 프로세서(126)는 또한, 전향 카메라(들)(112)로부터 주변 환경의 실제 객체들로부터 유래하는 비디오 데이터를 획득하도록 구성되며, 이는 가상 객체들로부터 유래하는 비디오 데이터와 함께, 아래에서 추가로 설명될 바와 같이 후속적으로 레코딩될 것이다.
- [0058] [0081] 유사하게, 오디오 프로세서(128)는 3-차원 데이터베이스(124)로부터 가상 객체들과 연관된 오디오 콘텐츠 및 메타 데이터를 획득하고 머리/객체 추적 서브시스템(120)으로부터 (아래에서 추가로 상세히 설명될 바와 같이, 최종 사용자(50)의 머리(54)에 대해 오디오에 대한 절대 메타 데이터를 로컬화하는 데 사용될) 최종 사용자(50)의 머리 포즈 데이터를 획득하고, 그로부터 공간화된 오디오를 렌더링하도록 구성되며, 이 공간화된 오디오는 그 후, 주변 환경의 실제 객체들로부터 유래하는 사운드들과 혼합되는 공간화된 사운드로의 변환을 위해 스피커(들)(108)에 전달된다.
- [0059] [0082] 오디오 프로세서(128)는 또한, 마이크로폰 조립체(116)로부터 주변 환경의 선택된 실제 객체(들)로부터만 유래하는 오디오 데이터를 획득하도록 구성되며, 이 오디오 데이터는 선택된 가상 객체들로부터의 공간화된 오디오 데이터와 함께, 각각의 가상 객체에 대한 최종 사용자(50)의 머리(54)에 대해 로컬화된 임의의 결과적인 메타 데이터(예컨대, 포지션, 배향 및 볼륨 데이터)뿐만 아니라 글로벌 메타 데이터(예컨대, 증강 현실 시스템



(100) 또는 최종 사용자(50)에 의해 포괄적으로 세팅된 볼륨 데이터)와 함께, 아래에서 추가로 설명될 바와 같이 후속적으로 레코딩될 것이다.

[0060] [0083] 증강 현실 시스템(100)은, 메모리(130), 메모리(130)에 비디오 및 오디오를 저장하도록 구성된 레코더(132), 및 최종 사용자(50) 또는 다른 최종 사용자들에 대한 후속 플레이백을 위해 메모리(130)로부터 비디오 및 오디오를 리트리브하도록 구성된 플레이어(134)를 더 포함한다. 레코더(132)는 오디오 프로세서(128)로부터, 선택된 가상 및 실제 객체들에 대응하는 공간화된 오디오 데이터(오디오 콘텐츠 오디오 데이터 및 메타 데이터 둘 모두)를 획득하고, 이 오디오 데이터를 메모리(130)에 저장하고, 선택된 가상 및 실제 객체들과 일치하는 가상 및 실제 객체들에 대응하는 비디오 데이터(비디오 콘텐츠 데이터 및 메타 데이터 둘 모두)를 추가로 획득한다. 플레이어(134)는 레코더(132) 및 메모리(130)가 로케이팅된 동일한 AR 시스템(100)에 로케이팅되는 것으로 예상되지만, 플레이어는 제3 자 AR 시스템에 또는 심지어, AR 시스템(100)에 의해 이전에 레코딩된 비디오 및 오디오를 플레이백하는 스마트폰 또는 컴퓨터 상에 로케이팅될 수 있다는 것이 인지되어야 한다.

[0061] [0084] 비디오 프로세서(126), 오디오 프로세서(128), 레코더(132) 및 플레이어(134)의 기능들을 수행하는 제어 서브시스템은 다양한 형태들 중 임의의 형태를 취할 수 있고, 다수의 제어기들, 예컨대, 하나 이상의 마이크로제어기들, 마이크로프로세서들 또는 CPU(central processing unit)들, 디지털 신호 프로세서들, GPU(graphics processing unit)들, 다른 집적 회로 제어기들, 이를테면, ASIC(application specific integrated circuit)들, PGA(programmable gate array)들, 예컨대, FPGA(field PGA)들 및/또는 PLU(programmable logic controller)들을 포함할 수 있다.

[0062] [0085] 비디오 프로세서(126), 오디오 프로세서(128), 레코더(132) 및 플레이어(134)의 기능들은 단일 통합 디바이스들에 의해 각각 수행될 수 있고, 비디오 프로세서(126), 오디오 프로세서(128), 레코더(132) 및/또는 플레이어(134)의 기능들 중 적어도 일부는 단일 통합 디바이스 내로 결합될 수 있거나, 또는 비디오 프로세서(126), 오디오 프로세서(128), 레코더(132) 또는 플레이어(134) 각각의 기능들은 여러 디바이스들 사이에서 분산될 수 있다. 예컨대, 비디오 프로세서(126)는 3-차원 데이터베이스(124)로부터 가상 객체들의 비디오 데이터를 획득하고 그로부터 합성 비디오 프레임들을 렌더링하는 GPU(graphics processing unit) 및 전향 카메라(들)(112)로부터 실제 객체들의 비디오 프레임들을 획득하는 CPU(central processing unit)를 포함할 수 있다. 유사하게, 오디오 프로세서(128)는 마이크로폰 조립체(116) 및 사용자 마이크로폰(122)으로부터 획득된 오디오 데이터를 프로세싱하는 DSP(digital signal processor) 및 3-차원 데이터베이스(124)로부터 획득된 오디오 데이터를 프로세싱하는 CPU를 포함할 수 있다. 레코더(132)의 레코딩 기능들 및 플레이어(134)의 플레이백 기능들은 CPU에 의해 수행될 수 있다.

[0063] [0086] 또한, 증강 현실 시스템(100)의 다양한 프로세싱 컴포넌트들은 분산된 서브시스템에 물리적으로 포함될 수 있다. 예컨대, 도 10a 내지 10d에 예시된 바와 같이, 증강 현실 시스템(100)은 이를테면, 유선 리드 또는 무선 연결(152)에 의해 최종 사용자(50)의 머리(54)에 장착된 컴포넌트들(예컨대, 디스플레이 서브시스템(102)의 프로젝션 서브시스템, 마이크로폰 조립체(116), 스피커들(104) 및 카메라들(114, 118))에 동작 가능하게 커플링된 로컬 프로세싱 및 데이터 모듈(150)을 포함한다. 로컬 프로세싱 및 데이터 모듈(150)은 다양한 구성들로 장착되는데, 이를테면, 프레임 구조(106)에 고정적으로 부착되고(도 10a), 헬멧 또는 모자(106a)에 고정적으로 부착되고(도 10b), 헤드폰들에 매립되고, 최종 사용자(50)의 몸통(58)에 제거 가능하게 부착되거나(도 10c), 또는 벨트-커플링 스타일 구성으로 최종 사용자(50)의 엉덩이(59)에 제거 가능하게 부착될 수 있다(도 10d). 증강 현실 시스템(100)은 이를테면, 유선 리드 또는 무선 연결성(158, 160)에 의해, 로컬 프로세싱 및 데이터 모듈(150)에 동작 가능하게 커플링되는 원격 프로세싱 모듈(154) 및 원격 데이터 리포지토리(156)를 더 포함하여서, 이들 원격 모듈들(154, 156)은 서로 동작 가능하게 커플링되고 로컬 프로세싱 및 데이터 모듈(150)에 대한 자원으로서 이용 가능하다.

[0064] [0087] 로컬 프로세싱 및 데이터 모듈(150)은 전력-효율적 프로세서 또는 제어기뿐만 아니라, 디지털 메모리 이를테면, 플래시 메모리를 포함할 수 있으며, 이들 둘 모두는 센서들로부터 캡처되고 그리고/또는 원격 프로세싱 모듈(154) 및/또는 원격 데이터 리포지토리(156)를 이용하여 취득 및/또는 프로세싱되고, 가능하다면, 이러한 프로세싱 또는 리트리브 후에 디스플레이 서브시스템(102)으로의 전달을 위한 데이터의 프로세싱, 캐싱(caching) 및 저장을 보조하기 위해 활용될 수 있다. 원격 프로세싱 모듈(154)은 데이터 및/또는 이미지 정보를 분석 및 프로세싱하도록 구성된 하나 이상의 비교적 강력한 프로세서들 또는 제어기들을 포함할 수 있다. 원격 데이터 리포지토리(156)는 "클라우드" 자원 구성에서 인터넷 또는 다른 네트워킹 구성을 통하여 이용 가능할 수 있는 비교적 대형-스케일 디지털 데이터 저장 설비를 포함할 수 있다. 일 실시예에서, 모든 데이터는 저

장되고 모든 컴퓨테이션은 로컬 프로세싱 및 데이터 모듈(150)에서 수행되어, 임의의 원격 모듈들로부터 완전히 자율적인 사용을 허용한다.

[0065] [0088] 위에서 설명된 다양한 컴포넌트들 사이의 커플링(coupling)들(152, 158, 160)은 유선 또는 광학 통신들을 제공하기 위한 하나 이상의 유선 인터페이스들 또는 포트들, 또는 이를테면, RF, 마이크로파 및 IR을 통해, 무선 통신들을 제공하기 위한 하나 이상의 무선 인터페이스들 또는 포트들을 포함할 수 있다. 일부 구현들에서, 모든 통신들은 유선일 수 있지만, 다른 구현들에서, 모든 통신들은 디스플레이 서브시스템(102)에서 사용되는 광섬유(들)를 제외하면, 무선일 수 있다. 또 다른 구현들에서, 유선 및 무선 통신들의 선택이 도 10a 내지 도 10d에 도시된 것과 상이할 수 있다. 따라서, 유선 또는 무선 통신들의 특정 선택이 제한되는 것으로 간주되어서는 안 된다.

[0066] [0089] 예시된 실시예에서, 디스플레이 서브시스템(102)의 광원(들) 및 구동 전자 장치(도시되지 않음), 및 머리/객체 추적 서브시스템(120) 및 객체 선택 디바이스(110)의 프로세싱 컴포넌트들, 및 오디오 프로세서(128)의 DSP는 로컬 프로세싱 및 데이터 모듈(150)에 포함될 수 있다. 비디오 프로세서(126)의 GPU 및 비디오 프로세서(126)의 CPU 및 오디오 프로세서(128)는 원격 프로세싱 모듈(154)에 포함될 수 있지만, 대안적인 실시예들에서, 이들 컴포넌트들 또는 그의 부분들은 로컬 프로세싱 및 데이터 모듈(150)에 포함될 수 있다. 3-차원 데이터베이스(124) 및 메모리(130)는 원격 데이터 리포지토리(156)와 연관될 수 있다.

[0067] [0090] 도 3에 예시된 오디오 프로세서(128)는 최종 사용자(50)에 의해 선택된 가상 및 실제 객체들로부터의 오디오 데이터를 프로세싱 및 레코딩하는데 있어 추가로 상세히 설명될 것이다. 도 11에 도시된 예시적인 시나리오에서, 최종 사용자(50)(예컨대, 부모)는 가상 드러머(V2) 객체, 실제 보컬리스트(R2), 예컨대 어린이, 가상 기타리스트(V3) 및 가상 베이스 기타리스트(V4)를 포함하는 4-인조 밴드로부터의 사운드들을 레코딩하기를 원하고, 가상 텔레비전(V1)으로부터의 사운드들을 레코딩함 없이 가상 텔레비전 상에서 뉴스 또는 스포츠를 모니터링하기를 원하고, 추가로 실제 부엌(R1), 예컨대 요리하는 사람으로부터의 사운드들을 레코딩하기를 원하지 않는다.

[0068] [0091] 도 12에 예시된 실시예에서, 오디오 프로세서(128)의 기능들은 가상 객체들로부터 유래하는 오디오를 프로세싱하는 CPU(180)와, 실제 객체들로부터 유래하는 오디오를 프로세싱하는 DSP(182) 사이에 분산된다. CPU(180)는 개별 가상 객체들(V1-Vn)에 대응하는 공간화된 오디오 데이터(EFX-V1-EFX-Vn)를 생성하도록 구성된 하나 이상의 특수 효과 모듈들(184)(이 경우에, 특수 효과 모듈들(1-n))을 포함한다. 이를 위해, 특수 효과 모듈들(184)은 3D 데이터베이스(124)로부터 가상 객체들(V1-Vn)에 대응하는 오디오 콘텐츠 데이터(AUD-V1 내지 AUD-Vn) 및 절대 메타 데이터(MD<sub>a</sub>-V1 내지 MD<sub>a</sub>-Vn) 뿐만 아니라 머리/객체 추적 서브시스템(120)으로부터 머리 포즈 데이터를 획득하고, 머리 포즈 데이터에 기초하여 최종 사용자(50)의 머리(54)에 대해 절대 메타 데이터(MD<sub>a</sub>-V1 내지 MD<sub>a</sub>-Vn)를 로컬화하고, 로컬화된 메타 데이터(예컨대, 포지션, 배향 및 볼륨 데이터)를 오디오 콘텐츠 데이터에 적용하여 가상 객체들(V1-Vn)에 대한 공간화된 오디오 데이터를 생성한다.

[0069] [0092] CPU(180)는, 혼합된 오디오 데이터(EFX)를 획득하기 위해 각각의 특수 효과 모듈(들)(184)로부터 수신된 공간화된 오디오 데이터(EFX-V1-EFX-Vn)를 혼합하도록 구성된 믹서(186), 및 다수의 사운드 채널들을 통해 스피커들(108)에 출력되는 최종 공간화된 오디오(AUD-OUT EFX)를 획득하기 위해 혼합된 공간화된 오디오 데이터에 글로벌 메타 데이터(MD-OUT)(예컨대, 글로벌 볼륨)를 적용하도록 구성된 글로벌 특수 효과 모듈(188)을 더 포함한다.

[0070] [0093] 중요하게, 특수 효과 모듈(들)(184)은 객체 선택 디바이스(110)를 통해 최종 사용자(50)에 의해 선택된 가상 객체들로부터 유래하는 오디오 콘텐츠 데이터 및 이들 선택된 가상 객체들에 대응하는 (로컬화된 그리고/또는 절대) 메타 데이터를, 메모리(130)(도 2에 도시됨)에의 저장을 위해 레코더(132)에 전송하도록 구성되고, 글로벌 특수 효과 모듈(188)은 메모리(130)에의 저장을 위해 레코더(132)에 글로벌 메타 데이터(MD-OUT)를 전송하도록 구성된다. 예시적인 실시예에서, 가상 오디오 콘텐츠 데이터(AUD-V2(즉, 가상 드러머), AUD-V3(즉, 가상 기타리스트), AUD-V4(즉, 가상 베이스 기타리스트))가 레코딩을 위해 선택되는 반면, 오디오 콘텐츠 데이터(AUD-V1)(즉, 가상 텔레비전)는 레코딩을 위해 선택되지 않는다. 따라서, 오디오 콘텐츠 데이터(AUD-V2, AUD-V3 및 AUD-V4) 및 대응하는 로컬화된 메타 데이터(MD-V2, MD-V3 및 MD-V4)는 도 13에 도시된 바와 같이 메모리(130)에 저장된다.

[0071] [0094] 대안적인 실시예에서, 선택된 가상 객체들로부터의 오디오 콘텐츠 데이터 및 대응하는 로컬화된/절대 메타 데이터 및 글로벌 메타 데이터를 메모리(130) 내에 개별적으로 저장하는 대신에 또는 그에 부가하여,

CPU(180)는 선택된 가상 객체들(AUD-V2, AUD-V3 및 AUD-V4)에만 대응하는 공간화된 오디오 데이터(EFX-V2, EFX-V3, EFX-V4)를 추가적으로 혼합하고 이 혼합된 공간화된 오디오 데이터에 글로벌 메타 데이터(MD-OUT)를 적용하여 선택된 가상 객체들(AUD-V2, AUD-V3 및 AUD-V4)로부터의 오디오만을 포함하는 공간화된 오디오를 획득함으로써 생성된 공간화된 오디오를 출력한다. 그러나, 이 경우에, 추가적인 오디오 혼합 기능이 CPU(180)에 통합될 필요가 있을 것이다.

[0072] [0095] DSP(182)는 마이크로폰 조립체(116)로부터 획득된 오디오 신호들을 프로세싱하고 특정 방향으로부터 그리고 이 경우에, 객체 선택 디바이스(110)를 통해 최종 사용자(50)에 의해 선택된 각각의 실제 객체의 방향으로부터 마이크로폰 조립체(116)에 의해 수신된 사운드들을 우선적으로 표현하는 오디오 신호들을 출력하도록 구성된다. 실제 객체의 포지션 및/또는 배향은 최종 사용자(50)의 머리(54)에 대해 이동할 수 있기 때문에, 실제 객체 추적 데이터가 머리/객체 추적 서브시스템(120)으로부터 수신될 수 있어서, 최종 사용자(50)의 머리(54)에 대한 실제 객체의 포지션 및/또는 배향의 임의의 변화가 고려될 수 있으므로, DSP(182)는 비교적 움직이는 실제 객체의 방향으로부터 마이크로폰 조립체(116)에 의해 수신된 사운드들을 우선적으로 표현하도록 오디오 출력을 동적으로 수정할 수 있다. 예컨대, 최종 사용자(50)는, 실제 객체가 선택되었을 때의 머리(54)의 배향에 대해 반시계 방향으로 90도만큼 자신의 머리(54)를 움직이는 경우, DSP(182)로부터 출력되는 오디오의 우선 방향은 시계 방향으로 90도만큼 동적으로 시프트될 수 있다.

[0073] [0096] 도 14를 참조하면, 마이크로폰 조립체(116)의 마이크로폰 엘리먼트들(118)은 마이크로폰 엘리먼트들(이 경우에, 마이크로폰 엘리먼트들(M1-Mn))의 위상화된 어레이의 형태를 취하며, 이들 각각은 주변 사운드 신호들을 검출하여 오디오 신호로 변환하도록 구성된다. 예시된 실시예에서, 마이크로폰 엘리먼트(118)는 본질적으로 디지털이고, 이에 따라 주변 사운드 신호를 디지털 오디오 신호들 그리고, 이 경우에 PDM(pulse density modulation) 신호들로 변환한다. 바람직하게는, 마이크로폰 엘리먼트들(118)은 오디오 출력의 방향성을 최대화하기 위해 서로 이격된다. 예컨대, 도 2에 도시된 바와 같이, 2개의 마이크로폰 엘리먼트들(118)은 프레임 구조(106)의 각각의 아암(arm)에 장착될 수 있지만, 2개 초과, 이를테면, 4개의 마이크로폰 엘리먼트들(118)이 프레임 구조(106)의 각각의 아암에 장착될 수 있다. 대안적으로, 프레임 구조(106)는 마이크로폰 엘리먼트들(118)이 프레임 구조(106)의 전방 및 후방에 장착될 수 있도록 설계될 수 있다. 이러한 방식으로, 마이크로폰 엘리먼트들(118)의 어레이는 사운드의 잠재적인 소스들의 모든 방향들을 커버하기 위해 최종 사용자(50)의 머리(54)를 둘러쌀 수 있다.

[0074] [0097] 마이크로폰 조립체(116)는, 대응하는 마이크로폰 엘리먼트들(118)로부터 각각의 디지털 오디오 신호들을 각각 수신하고, PDM 포맷으로부터 보다 쉽게 조작 가능한 PCM(pulse code modulation)으로 디지털 오디오 신호들을 변환하기 위해 "데시메이션(decimation)"으로서 지칭되는 디지털 필터 동작을 수행하도록 구성된 복수의 DMIC(digital microphone interface)들(190)(이 경우에, 각각의 마이크로폰 엘리먼트(M)마다 하나씩의 DMIC1-DMICn)을 더 포함한다. DMIC들(190) 각각은 또한 디지털 오디오 신호들에 대해 고정 이득 제어를 수행한다.

[0075] [0098] DSP(182)는 복수의 오디오 프로세싱 모듈들(200)을 포함하며, 이들 각각은 마이크로폰 조립체(116)에 의해 출력된 디지털 오디오 신호를 프로세싱하고, 선택된 실제 객체(R1 내지 Rm 중 하나)의 방향에서 마이크로폰 조립체(116)에 의해 수신된 사운드를 우선적으로 표현하는 방향성 오디오 신호(AUD-R)(방향성 오디오 신호들(AUD-R1 내지 AUD-Rm) 중 하나)를 출력하도록 구성된다. 각각의 오디오 프로세싱 모듈들(200)에 의해 출력되는 방향성 오디오 신호들(AUD-R1 내지 AUD-Rm)은 선택된 모든 실제 객체들로부터 유래하는 사운드를 우선적으로 표현하는 방향성 오디오 출력(AUD-OUT MIC)으로 결합된다. 예시된 실시예에서, DSP(182)는 객체 선택 디바이스(110)를 통해 최종 사용자(50)에 의해 선택된 각각의 실제 객체에 대한 오디오 프로세싱 모듈(200)의 하나의 인스턴스를 생성한다.

[0076] [0099] 이를 위해, 오디오 프로세싱 모듈들(200) 각각은 복수의 지연 엘리먼트들(194)(이 경우에, 각각의 마이크로폰 엘리먼트(M)마다 하나씩의 지연 엘리먼트들(D1-Dn)), 복수의 이득 엘리먼트들(196)(이 경우에, 각각의 마이크로폰 엘리먼트(M)마다 하나씩의 이득 엘리먼트들(G1-Gn)) 및 합산기(198)의 형태의 프로세싱 파라미터들을 포함한다. 지연 엘리먼트들(194)은 마이크로폰 조립체(116)의 대응하는 이득 증폭기들(192)로부터 수신된 증폭된 디지털 신호들에 지연 팩터들을 각각 적용하고, 이득 엘리먼트들(196)은 지연된 디지털 신호들에 이득 팩터들을 각각 적용한다. 합산기(198)(S)는 각각의 방향성 오디오 신호(AUD-R)를 각각 생성하기 위해 이득 조정 및 지연된 신호들을 더한다.

[0077] [0100] 마이크로폰 엘리먼트들(118)은 공간적으로 배열되고 각각의 오디오 프로세싱 모듈(200)의 지연 엘리먼트들(194) 및 이득 엘리먼트들(196)은, 방향성 극 패턴에 따라 주변 사운드를 수신하게 하는 방식으로 마이크로



폰 조립체(116)로부터 수신된 디지털 오디오 신호들에 적용된다(즉, 특정 각 방향 또는 방향들로부터 도달하는 사운드들은 다른 각 방향들로부터 도달하는 사운드들보다 더 강조될 것임). DSP(182)는 지연 엘리먼트들(194)의 지연 팩터들 및 이득 엘리먼트들(196)의 이득 팩터들을 변경함으로써 방향성 오디오 신호들(AUD-R1 내지 AUD-Rm)의 방향성 및 이에 따라, 결합된 방향성 오디오 출력(AUD-OUT MIC)을 수정하도록 구성된다.

[0078] [00101] 따라서, 오디오 출력(AUD-OUT MIC)의 방향성은 선택된 실제 객체에 기초하여 수정되는데, 예컨대 사운드가 우선적으로 수신되는 방향 또는 방향들은 선택된 실제 객체 또는 소스들의 방향을 따라 세팅될 수 있다는 것이 인지될 수 있다.

[0079] [00102] 예컨대, 도 15a를 참조하면, 각각 2개의 특정 방향들( $D_a$  및  $D_b$ )을 따라 2개의 실제 객체들( $R_a$  및  $R_b$ )이 선택되는 경우, DSP(182)는 오디오 프로세싱 모듈들(200)의 2개의 인스턴스들을 생성하고, 이들 오디오 프로세싱 모듈들(200) 각각 내에서, 실제 객체들( $R_a$  및  $R_b$ )의 방향들( $D_a$  및  $D_b$ )과 정렬된 2개의 로브(lobe)들을 갖는 수신 이득 패턴이 생성되도록 각각의 오디오 프로세싱 모듈(200)의 모든 지연 엘리먼트들(194) 및 이득 엘리먼트들(196)에 대한 각각의 지연 팩터들 및 이득 팩터들을 선택할 것이다. 최종 사용자(50)의 머리(54)에 대한 실제 객체들( $R_a$  및  $R_b$ )의 배향이 변하는 경우, 실제 객체들( $R_a$  및  $R_b$ )의 특정 방향들은 변할 수 있으며, 이 경우에, DSP(182)는 도 15b에 예시된 바와 같이, 수신 이득 패턴이 방향들( $D_c$  및  $D_d$ )과 정렬된 2개의 로브들을 갖도록, 각각의 오디오 프로세싱 모듈(200)의 모든 지연 엘리먼트들(194) 및 이득 엘리먼트들(196)에 대해 상이한 지연 팩터들 및 이득 팩터들을 선택할 수 있다.

[0080] [00103] 오디오 출력(AUD-OUT MIC)의 방향성의 이러한 동적 수정을 용이하게 하기 위해, 지연/이득 값들의 상이한 세트들 및 대응하는 우선 방향들은 DSP(182)에 의한 액세스를 위해 메모리(130)에 저장될 수 있다. 즉, DSP(182)는 각각의 선택된 실제 객체(R)의 방향을 메모리(130)에 저장된 가장 근접한 방향 값과 매칭시키고, 그 선택된 방향에 대한 지연/이득 팩터들의 대응하는 세트를 선택한다.

[0081] [00104] 마이크로폰 엘리먼트들(118)이 디지털인 것으로 설명되지만, 마이크로폰 엘리먼트들(118)은 대안적으로 아날로그일 수 있다는 것이 주의되어야 한다. 또한, 지연 엘리먼트들(194), 이득 엘리먼트들(196) 및 합산기(198)가 DSP(182) 내에 상주하는 소프트웨어 컴포넌트들인 것으로 개시되고 예시되지만, 지연 엘리먼트들(194), 이득 엘리먼트들(196) 및 합산기(198) 중 임의의 하나 이상은, DSP(182) 외부에 상주하지만 그의 제어하에 있는 아날로그 하드웨어 컴포넌트들을 포함할 수 있다. 그러나, 소프트웨어-기반 오디오 프로세싱 모듈들(200)의 사용은 몇몇 별개의 실제 객체들로부터의 사운드가 우선적으로 수신되고 동시에 프로세싱될 수 있게 한다.

[0082] [00105] 도 12를 다시 참조하면, DSP(182)는 또한 사용자 마이크로폰(122)으로부터 음성 데이터를 수신하고, 이를 방향성 오디오 출력(AUD-OUT MIC)과 결합한다. 선택적인 실시예에서, DSP(182)는 가상 객체들로부터 유래한, 스피커들(108)로부터의 사운드들에 대해 AEC(acoustic echo cancellation) 및 NS(noise suppression) 기능들을 수행하도록 구성된다. 즉, 마이크로폰 조립체(116)는 사운드가 우선적으로 수신되는 방향이 스피커들(108)과 일치하지 않더라도 스피커(108)에 의해 방출된 사운드들을 감지할 수 있다. 이를 위해, 글로벌 특수 효과 모듈(188)에 의해 스피커들(108)로 출력된 공간화된 오디오 데이터가 또한 DSP(182)로 입력되며, 이는 공간화된 오디오 데이터를 사용하여 마이크로폰 조립체(116) 내로 스피커들(108)에 의해 출력되는 결과적인 사운드들(노이즈로 간주됨)을 억제하고 마이크로폰 조립체(116) 내로 스피커들(108)로부터의 피드백으로부터 발생하는 임의의 에코들을 소거한다.

[0083] [00106] 중요하게, DSP(182)는 추가로, 오디오 콘텐츠 데이터로서 메모리(130)(도 2에 도시됨)에의 저장을 위해, 방향성 오디오 출력(AUD-OUT MIC) 및 로컬화된 메타 데이터(예컨대, 방향성 오디오 출력(AUD-OUT MIC)이 유래한 실제 객체의 위치 및 배향)을 레코더(132)에 전송하도록 구성된다. 도 11에 예시된 예시적인 실시예에서, 로컬화된 메타 데이터는 실제 객체(AUD-R2)(즉, 실제 보컬리스트)에 대응한다. 따라서, 방향성 오디오 출력(AUD-OUT MIC)(이는 실제 객체(AUD-R2)에 우선적으로 대응함) 및 대응하는 로컬화된 메타 데이터(MD-R2)는 도 13에 도시된 바와 같이 메모리(130)에 저장된다.

[0084] [00107] 선택적인 실시예에서, (공간화될 수 있는) 방향성 오디오 출력(AUD-OUT MIC)은 최종 사용자(50)에게의 플레이백을 위해 스피커들(108) 또는 다른 스피커들에 입력될 수 있다. 방향성 오디오 출력(AUD-OUT MIC)은, 사운드들의 선명도 또는 사실감에 영향을 주도록, 가상 소스들로부터 유래하는 공간화된 오디오 데이터와 동일한 방식으로 공간화되어 사운드들이 실제 객체들의 위치들로부터 유래하는 것처럼 최종 사용자(50)에게 나타나게 할 수 있다. 즉, 로컬화된 메타 데이터(예컨대, 방향성 오디오 출력(AUD-OUT MIC)이 우선적으로 유래한 실제 객체의 위치 및 배향)은 공간화된 오디오 데이터를 획득하기 위해 방향성 오디오 출력(AUD-OUT MIC)에 적용

될 수 있다.

- [0085] [00108] 다른 선택적인 실시예에서, 최종 사용자(50)에 의해 선택된 실제 객체 또는 심지어 가상 객체로부터 유래하는 사운드가 프로파일링될 수 있다. 특히, DSP(182)는 타겟 사운드의 유형을 결정하기 위해, 선택된 객체로부터의 사운드의 특성을 분석하여 다른 실제 객체들로부터 유래하는 사운드들의 특성들과 비교할 수 있다. 그 후, DSP(182)는 원하는 경우, 메모리(130)(도 2에 도시됨) 내로의 레코더(132)에 의한 레코딩을 위해 방향성 오디오 출력(AUD-OUT MIC)에, 이들 실제 객체들로부터 유래하는 모든 오디오 데이터를 포함시킬 수 있다. 예컨대, 최종 사용자(50)가 음악 객체들(AUD-V2, AUD-V3, AUD-V4, AUD-R2) 중 임의의 것을 선택한 경우, DSP(182)는 모든 음악적 실제 객체들을 우선적으로 감지하도록 마이크로폰 조립체(116)를 제어할 수 있다.
- [0086] [00109] 예시된 실시예에서, DSP(182)는, (머리/객체 추적 서브시스템(120)으로부터 수신된 실제 객체 추적 데이터에 의해 표시된 바와 같이) 최종 사용자(50)에 의해 선택된 실제 객체(198)가 디스플레이 서브시스템(102)의 시야 밖으로 이동하는 경우조차도, 메모리(130)에의 레코딩을 위해 방향성 오디오 출력(AUD-OUT MIC)을 레코더(130)에 계속 출력한다. 대안적인 실시예에서, DSP(182)는 최종 사용자(50)에 의해 선택된 실제 객체(198)가 디스플레이 서브시스템(102)의 시야 밖으로 이동하자마자, 메모리(130)에의 레코딩을 위해 방향성 오디오 출력(AUD-OUT MIC)을 레코더(130)에 출력하기를 중단하고, 최종 사용자(50)에 의해 선택된 실제 객체(198)가 디스플레이 서브시스템(102)의 시야 내로 다시 이동하자마자, 메모리(130)에의 레코딩을 위해 레코더(130)로의 방향성 오디오 출력(AUD-OUT MIC)의 출력을 다시 개시한다.
- [0087] [00110] 오디오 프로세서(128)(예시된 실시예에서, CPU(180) 및 DSP(182))가 메모리(130)에의 저장을 위해, 선택된 가상 객체들 및 실제 객체들로부터 유래하는 오디오 콘텐츠 데이터(예시적인 경우에, 오디오 콘텐츠 데이터(AUD-V2, AUD-V3, 및 AUD-V4, 및 AUD-MIC)) 및 로컬화된 메타 데이터(예시적인 경우에, MD-V2, MD-V3, MD-V4, 및 MD-R2) 및 글로벌 메타 데이터(MD-OUT)를 레코더(132)에 전송하는 것과 유사한 방식으로, 비디오 프로세서(126)는 도 13에 예시된 바와 같이 가상 객체들 및 실제 객체들로부터 유래하는 비디오 콘텐츠 데이터(예시적인 경우에, 비디오 콘텐츠 데이터(VID-V2, VID-V3, VID-V4 및 VID-R2))를 전송할 수 있다. 가상 객체들의 경우에, 비디오 프로세서(126)는 추가의 프로세싱 없이 3D 데이터베이스(124)로부터 가상 객체들을 단순히 획득하고 메모리(130)에의 저장을 위해 이들 가상 객체들을 레코더(132)에 전송한다. 실제 객체들의 경우에, 비디오 프로세서(126)는 카메라(들)(112)로부터 획득된 비디오로부터, 선택된 실제 객체들 중 임의의 것을 추출 또는 "컷 오프(cut off)"하고, 이러한 실제 객체들을 가상 객체들로서 메모리(130)에 저장할 수 있다. 도 11에 예시된 예시적인 경우에, 실제 보컬리스트(R2)에 대한 비디오는 가상 객체(VID-R2)로서 레코딩될 수 있다. 선택적인 실시예에서, 비디오 프로세서(126)는 메모리(130)에의 저장을 위해, 카메라(들)(112)로부터 획득된 전체 비디오(비-선택 가상 및 실제 객체들에 대응하는 비디오를 포함함)를 레코더(132)에 전송한다.
- [0088] [00111] 플레이어(134)는 메모리(130) 내에 레코딩된 비디오 및/또는 오디오를, 비디오/오디오를 레코딩한 오리지널 최종 사용자(50) 또는 제3 자 사용자일 수 있는 플레이어백 사용자(50')(도 16a에 도시됨)에 플레이어백하도록 구성된다. 오디오/비디오는 플레이어백 사용자(50')에 의해 주어진 커맨드들, 예컨대, 사용자 마이크로폰(122)을 통한 음성 커맨드들에 대한 응답으로 플레이어(134)에 의해 선택적으로 플레이어백될 수 있다. 예컨대, 플레이어백 사용자(50')는 "virtual audio on/off" 커맨드를 이용하여 가상 오디오 플레이어백을 턴 온 또는 턴 오프시키거나 또는 "display on/off"를 사용하여 가상 비디오 플레이어백을 턴 온 또는 턴 오프시키거나, 또는 "real-audio on/off" 커맨드를 사용하여 실제 오디오 플레이어백을 턴 온 또는 턴 오프시킬 수 있다.
- [0089] [00112] 예시된 실시예에서, 오디오 프로세서(128)는 메모리(130)로부터 오디오 콘텐츠 데이터 및 메타 데이터(선택된 가상 및 실제 객체들에 대응함)를 리트리브하고, 오디오 콘텐츠 데이터 및 메타 데이터로부터 공간화된 오디오를 렌더링하고, 스피커들(108)을 통해 플레이어백 사용자(50')에게의 플레이어백을 위해 플레이어(134)에 공간화된 오디오를 전달한다. (콘텐츠 및 메타 데이터 대신에) 혼합된 공간화된 오디오 데이터가 저장되는 대안적인 실시예에서, 플레이어(134)는 오디오 데이터의 재-렌더링 또는 다른 방식의 추가의 프로세싱 없이 플레이어백 사용자(50')에게의 플레이어백을 위해 메모리(130)로부터 오디오 데이터를 단순히 획득할 수 있다.
- [0090] [00113] 또한, 예시된 실시예에서, 비디오 프로세서(126)는 (선택된 가상 및 실제 객체들에 대응하는) 비디오 콘텐츠 데이터 및 메타 데이터를 리트리브하고, 비디오 콘텐츠 데이터 및 메타 데이터로부터 비디오를 렌더링하고, 스피커들(108)을 통한 오디오의 플레이어백과 동기하여, 디스플레이 서브시스템(102)을 통한 플레이어백 사용자(50')에게의 플레이어백을 위해 플레이어(134)에 비디오를 전달한다. 선택적으로, 카메라(들)(112)에 의해 캡처된 모든 비디오 데이터가 저장되는 경우에, 플레이어(134)는 비디오 데이터의 렌더링 또는 다른 방식의 추가의 프로세싱 없이, 플레이어백 사용자(50')에게의 플레이어백을 위해 메모리(130)로부터 비디오 데이터를 단순히 획득



할 수 있다. 증강 현실 시스템(10)은 선택된 가상 및 실제 객체들에 대응하는 비디오만을 플레이백하거나 카메라(들)(112)에 의해 캡처된 전체 비디오를 플레이백하는 옵션을 플레이백 사용자(50')에게 제공할 수 있다.

[0091] [00114] 일 실시예에서, 플레이백 사용자(50')의 현재 머리 포즈는 비디오/오디오의 플레이백 동안 고려되지 않는다. 대신에, 비디오/오디오는 비디오/오디오 데이터의 레코딩 동안 원래 검출된 머리 포즈를 사용하여 플레이백 사용자(50')에게 플레이백되며, 이는 메모리(130) 내에 오디오/비디오 콘텐츠 데이터와 함께 저장된 로컬화된 메타 데이터에 반영될 것이거나, 또는 혼합된 공간화된 오디오가 메타 데이터 없이 레코딩된 경우, 머리 포즈는 메모리(130)에 저장된 혼합된 공간화된 오디오 내에 반영될 것이다. 이 경우에, 플레이백 사용자(50')는, 오리지널 최종 사용자(50)에 의해 선택된 가상 및 실제 객체들로부터 유래한 오디오만 그리고 선택적으로 비디오만이 플레이백될 것임을 제외하면, 오리지널 최종 사용자(50)가 비디오/오디오를 경험한 것과 동일한 방식으로 비디오/오디오를 경험할 것이다. 이 경우에, 플레이백 사용자(50')의 머리 포즈가 고려될 것이므로, 플레이백 사용자(50')는 증강 현실에 몰입되지 않을 수 있다. 오히려, 플레이백 사용자(50')는 헤드셋을 사용한 오디오 플레이백을 경험할 수 있거나(따라서 오디오는 환경에 의해 영향을 받지 않을 것임), 또는 플레이백 사용자(50')는 조용한 룸에서의 오디오 플레이백을 경험할 수 있다.

[0092] [00115] 대안적인 실시예에서, 플레이백 사용자(50')의 현재 머리 포즈는 비디오/오디오의 플레이백 동안 고려될 수 있다. 이 경우에, 비디오/오디오의 레코딩 동안 플레이백 사용자(50')의 머리 포즈는 메모리(130)에 비디오/오디오 콘텐츠 데이터와 함께 저장된 메타 데이터에 통합될 필요가 없는데, 그 이유는 플레이백 동안 검출된 플레이백 사용자(50')의 현재 머리 포즈가 비디오/오디오 데이터를 재-렌더링하는 데 사용될 것이기 때문이다. 대신, 메모리(130)에 저장된 절대 메타 데이터(예컨대, 3D 장면에서 이들 가상 객체들의 볼륨 및 절대 위치 및 방향뿐만 아니라, 가상 소스 부근의 임의의 가상 또는 실제 객체들을 포함한 각각의 가상 객체를 둘러싼 공간 음향, 룸 치수들, 벽/바닥 재료들 등)은 플레이백 사용자(50')의 현재 머리 포즈를 사용하여 플레이백 사용자(50')의 머리 포즈에 대해 로컬화되고, 그 후 오디오/비디오를 렌더링하는 데 사용될 것이다. 따라서, 플레이백 사용자(50')는 비디오/오디오의 플레이백 동안 증강 현실에 몰입될 것이다.

[0093] [00116] 플레이백 사용자(50')는 비디오/오디오가 레코딩된 오리지널 공간 환경(예컨대, "동일한 물리적 룸")에서 증강 현실을 경험할 수 있거나 또는 새로운 물리적 또는 가상 공간 환경(예컨대, "상이한 물리적 또는 가상 룸")에서 증강 현실을 경험할 수 있다.

[0094] [00117] 증강 현실이 비디오/오디오가 레코딩된 오리지널 공간 환경에서 플레이백 사용자(50')에 의해 경험되는 경우, 선택된 객체들과 연관된 절대 메타 데이터는 공간화된 오디오의 정확한 플레이백을 위해 수정될 필요는 없다. 반대로, 증강 현실이 새로운 공간 환경에서 플레이백 사용자(50')에 의해 경험되는 경우, 객체들과 연관된 절대 메타 데이터는 새로운 공간 환경에서 오디오/비디오의 정확한 렌더링을 위해 수정될 필요가 있을 수 있다.

[0095] [00118] 예컨대, 예시적인 실시예에서, 가상 객체들(AUD-V2(즉, 가상 드러머), AUD-V3(즉, 가상 기타리스트), AUD-V4(즉, 가상 베이스 기타리스트)) 및 실제 객체(즉, 실제 보컬리스트)로부터의 오디오/비디오 콘텐츠가 도 16a에 예시된 바와 같이 작은 룸(250)에서 레코딩될 수 있다. 가상 객체들(AUD-V2(즉, 가상 드러머), AUD-V3(즉, 가상 기타리스트), AUD-V4(즉, 가상 베이스 기타리스트)) 및 실제 객체(즉, 실제 보컬리스트)로부터 이전에 레코딩된 오디오는 도 16b에 예시된 바와 같이 콘서트 홀(252)에서 플레이백될 수 있다. 증강 현실 시스템(10)은 콘서트 홀(252) 내의 임의의 장소에 객체들을 채포지정할 수 있고, 콘서트 홀(252) 내의 각각의 객체의 새로운 포지션들뿐만 아니라 콘서트 홀(252)에서 각각의 객체를 둘러싸는 공간 음향을 포함하는 절대 메타 데이터가 생성되거나 그렇지 않으면 획득될 수 있다. 이 절대 메타 데이터는 그 후 플레이백 사용자(50')의 현재 머리 포즈를 사용하여 로컬화될 수 있고, 그 후 플레이백 사용자(50')에게의 플레이백을 위해 콘서트 홀(252)에서 오디오 및 비디오를 렌더링하는 데 사용될 수 있다.

[0096] [00119] 증강 현실 시스템(100)의 어레인지먼트 및 기능을 설명하였지만, 적어도 하나의 객체를 선택하고 이들 선택된 객체(들)로부터의 오디오 및 비디오를 레코딩하도록 증강 현실 시스템(100)을 사용하는 하나의 방법(300)이 이제 도 17과 관련하여 설명될 것이다.

[0097] [00120] 먼저, 최종 사용자(50)는 객체 선택 디바이스(110)를 통해 공간 환경에서 적어도 하나의 객체(예컨대, 실제 및/또는 가상)를 지속적으로 선택한다(단계 302). 객체(들)는, 예컨대, 최종 사용자(50)의 시야(60)에서 3-차원 커서(62)를 이동시키고, 3-차원 커서(62)(도 5에 도시됨)로 객체(들)를 선택함으로써, 최종 사용자(50)의 시야(60)에서 선택될 수 있다. 또는, 객체(들)는 손 제스처들(도 6 또는 도 7에 도시됨)을 사용하여 또는 음성 커맨드들을 사용하여 선택될 수 있다. 다수의 객체들은 개별적으로 선택될 수 있거나, 또는 예컨대, 객체

들 주위에 라인(64)을 그림으로써(도 8에 도시됨) 또는 최종 사용자(50)의 시야(60)의 각도 범위(66)(이는 최종 사용자(50)의 시야(60)의 전체 각도 범위보다 작을 수 있음)를 정의하고, 최종 사용자(50)의 시야(60)의 정의된 각도 범위(66) 내의 모든 객체를 선택함으로써(도 9에 도시됨) 포괄적으로 선택될 수 있다.

[0098] [00121] 다음으로, 공간 환경 내의 모든 가상 객체들에 대한 오디오 및 비디오 콘텐츠뿐만 아니라 가상 객체들과 연관된 절대 메타 데이터가 획득된다(단계 304). 다음으로, 최종 사용자(50)의 현재 머리 포즈가 추적되고(단계 306), 절대 메타 데이터는 현재 머리 포즈 데이터를 사용하여 최종 사용자(50)의 머리(54)에 대해 로컬화되고(단계 308), 가상 객체들의 오디오 및 비디오 콘텐츠에 적용되어 각각의 가상 객체들 모두에 대한 비디오 데이터 및 공간화된 오디오 데이터를 획득한다(단계 310). 3D 장면에서의 각각의 가상 객체들 모두에 대한 공간화된 오디오 데이터가 혼합되고(단계 312), 글로벌 메타 데이터는 혼합된 공간화된 오디오 데이터에 적용되어 3D 장면에서의 모든 가상 객체들에 대한 최종 공간화된 오디오를 획득하며(단계 314), 이는 그 후 최종 사용자(50)에 의한 지각을 위한 사운드로 변환된다(단계 316). 다음으로, 단계(310)에서 획득된 비디오 데이터는 최종 사용자(50)에 의한 지각을 위한 이미지 프레임들로 변환된다(단계 318). 다음으로, 단계(302)에서 최종 사용자(50)에 의해 선택된 모든 가상 객체(들)에 대한 오디오/비디오 콘텐츠 및 모든 연관된 메타 데이터(절대 및 로컬화된 메타 데이터 둘 모두)가 레코딩된다(단계 320).

[0099] [00122] 단계들(304-320)과 병렬로, 최종 사용자(50)의 머리(54)에 대한 선택된 실제 객체(들)의 포지션 및/또는 배향이 추적되고(단계 322), 선택된 실제 객체(들)로부터 유래하는 사운드가 실제 객체(들)의 추적된 포지션 및 배향에 기초하여 우선적으로 감지된다(단계 324). 다음으로, 선택된 실제 객체(들)의 이미지들이 캡처되고(단계 326), 선택적으로 가상 비디오 콘텐츠로 변환된다. 다음으로, 선택된 실제 객체(들)로부터의 우선적으로 감지된 사운드와 연관된 오디오 콘텐츠 및 선택된 실제 객체(들)의 캡처된 이미지들과 연관된 비디오 콘텐츠뿐만 아니라 모든 연관된 메타 데이터(실제 객체(들)의 위치 및 배향)가 선택된 실제 객체(들) 각각에 대해 레코딩된다(단계 328).

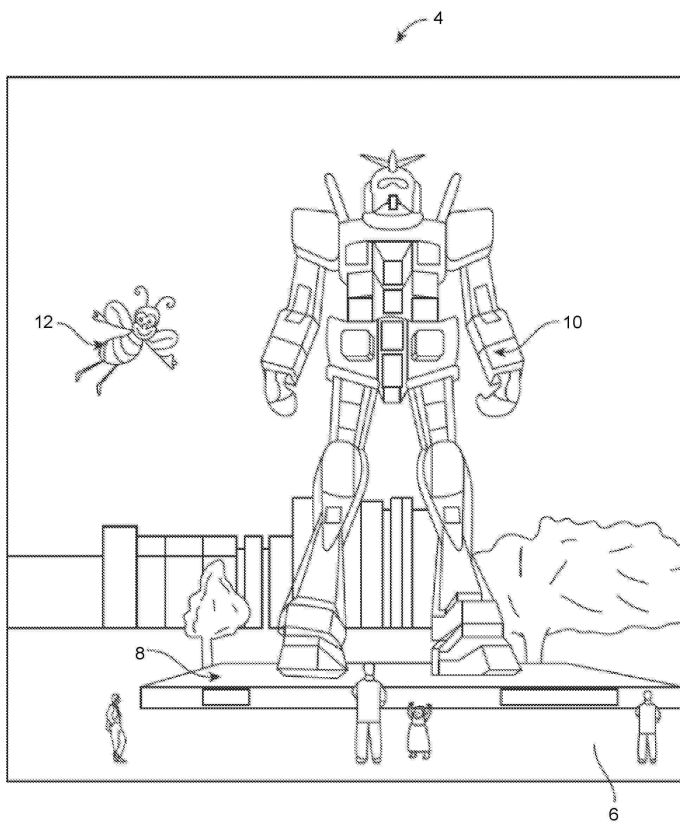
[0100] [00123] 적어도 하나의 객체의 이전에 레코딩된 오디오 및 비디오를 플레이백 사용자(50')에게 플레이백하기 위해 증강 현실 시스템(100)을 사용하는 하나의 방법(400)이 이제 도 18과 관련하여 설명될 것이다. 이러한 오디오 및 비디오는 도 17의 방법(300)에서 위에서 설명된 방식으로 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터로서 이전에 레코딩되었을 수 있다. 객체(들)는 실제 및/또는 가상일 수 있고, 최종 사용자(50)에 의해 지속적으로 선택되었을 수 있다. 예시적인 방법(400)에서, 오디오 및 비디오는 작은 룸(250)과 같은 오리지널 공간 환경에서 이전에 레코딩되었고, 도 16a 및 도 16b와 관련하여 설명된 바와 같이 콘서트 홀(252)과 같은, 오리지널 공간 환경과 상이한 새로운 공간 환경에서 플레이백된다.

[0101] [00124] 먼저, 이전에 레코딩된 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터가 획득된다(단계 402). 새로운 공간 환경이 적어도 부분적으로 가상인 경우, 새로운 공간 환경과 연관된 부가적인 가상 콘텐츠(오디오 또는 비디오)가 또한 획득될 수 있다. 그 후, 객체(들)는 새로운 공간 환경 내에 재포지셔닝되며, 이는 플레이백 사용자(50')로부터의 입력에 응답하여 이루어질 수 있다(단계 404). 그 후, 새로운 공간 환경에 포지셔닝된 객체(들)에 대응하는 절대 메타 데이터가 획득되고(단계 406), 플레이백 사용자(50')의 머리 포즈가 새로운 공간 환경에서 추적되고(단계 408), 절대 메타 데이터는 플레이백 사용자(50')의 추적된 머리 포즈에 기초하여 플레이백 사용자(50')에 대해 로컬화된다(단계 410). 다음으로, 오디오 및 비디오는, 새로운 공간 환경에서 로컬화된 메타 데이터에 기초하여, 리트리브된 오디오 콘텐츠 데이터 및 비디오 콘텐츠 데이터로부터 렌더링된다(단계 412). 렌더링된 오디오 및 비디오는 그 후, 플레이백 사용자(50')에 의한 동기식 지각을 위해 사운드 및 이미지 프레임들로 각각 변환된다(단계 414).

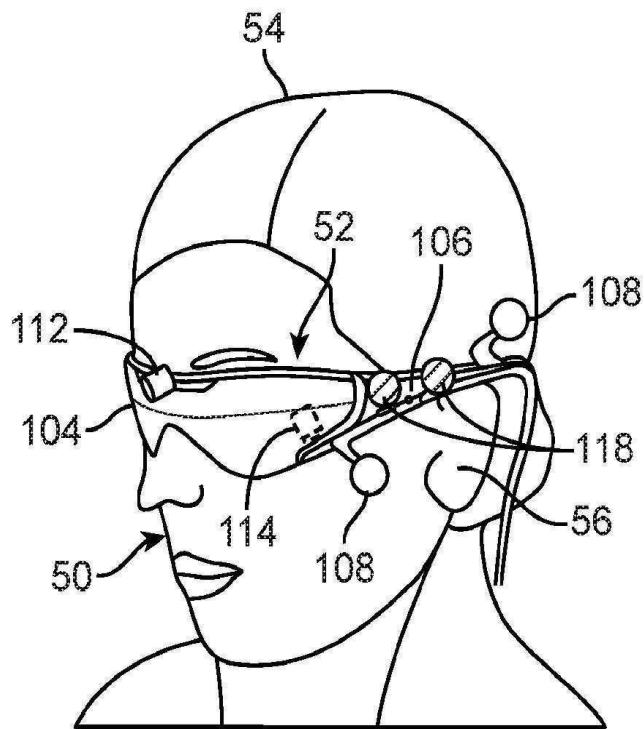
[0102] [00125] 위의 명세서에서, 본 발명은 본 발명의 특정 실시예들을 참조하여 설명되었다. 그러나, 본 발명의 더 넓은 사상 및 범위를 벗어나지 않으면서 다양한 수정들 및 변경들이 본 발명에 대해 행해질 수 있다는 것은 명백할 것이다. 예컨대, 위에서-설명된 프로세스 흐름들은, 프로세스 동작들의 특정한 순서를 참조하여 설명된다. 그러나, 설명된 프로세스 동작들 대부분의 순서는 본 발명의 범위 또는 동작에 영향을 주지 않으면서 변경될 수 있다. 따라서, 명세서 및 도면들은 제한적인 의미보다는 예시적인 의미로 간주될 것이다.

도면

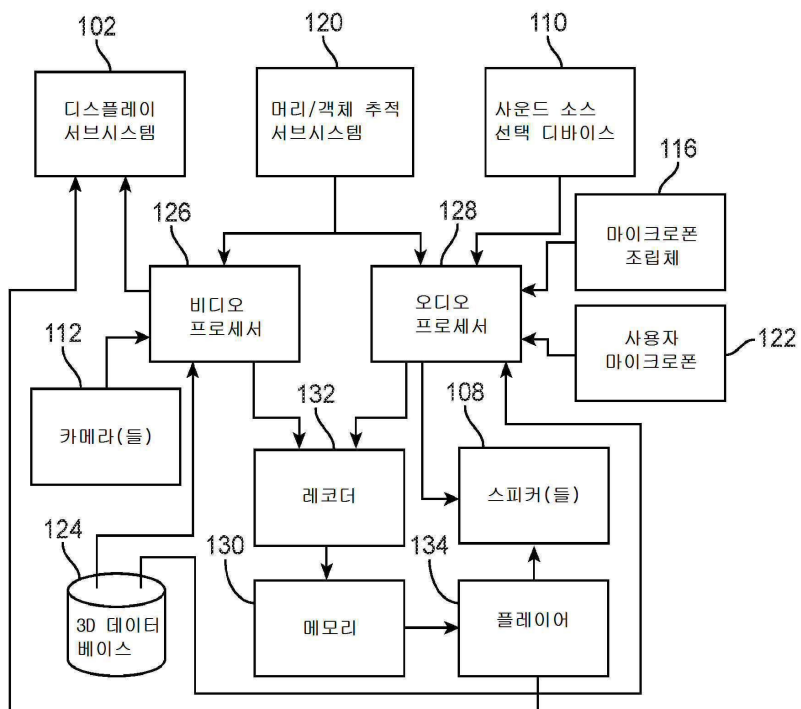
도면1



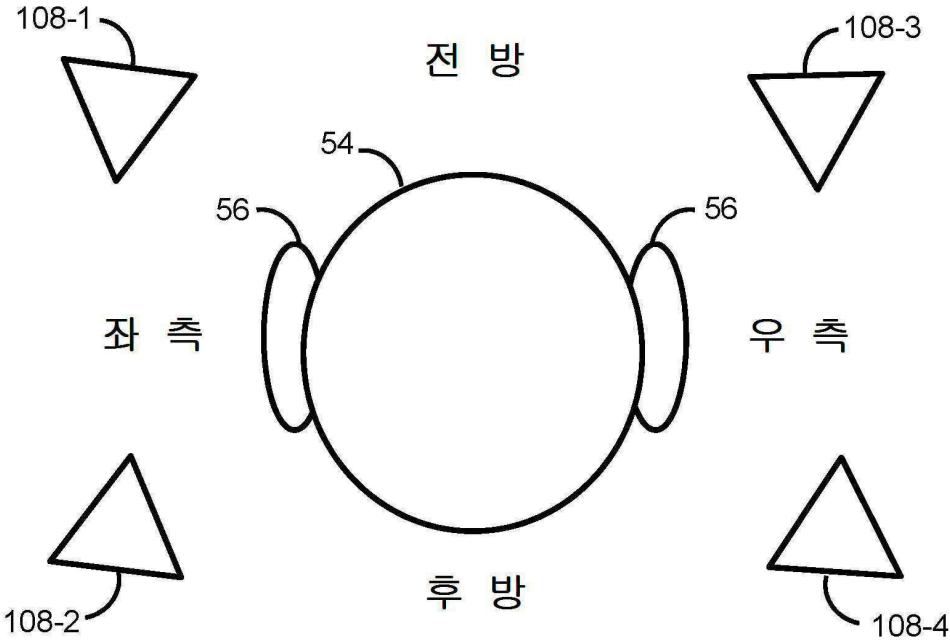
도면2



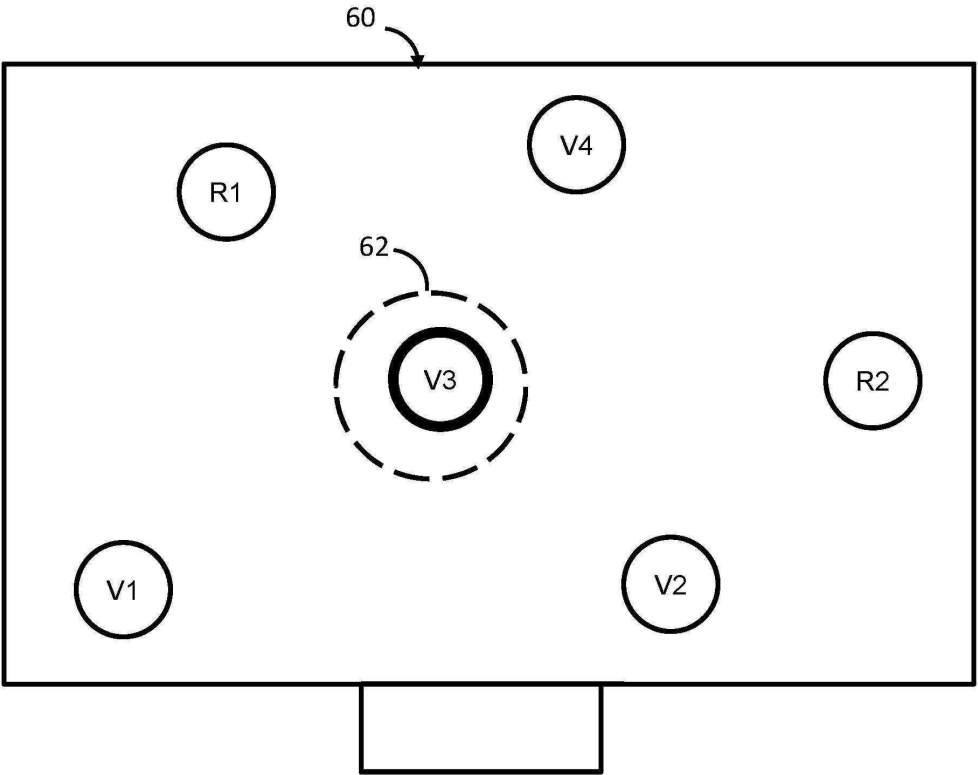
도면3



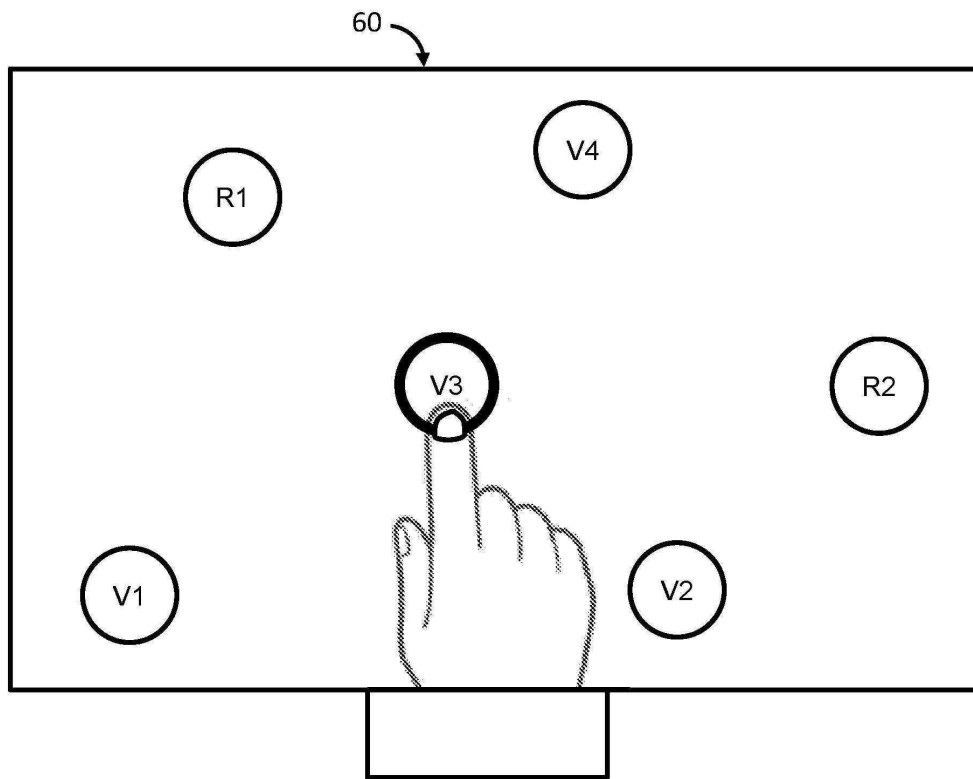
도면4



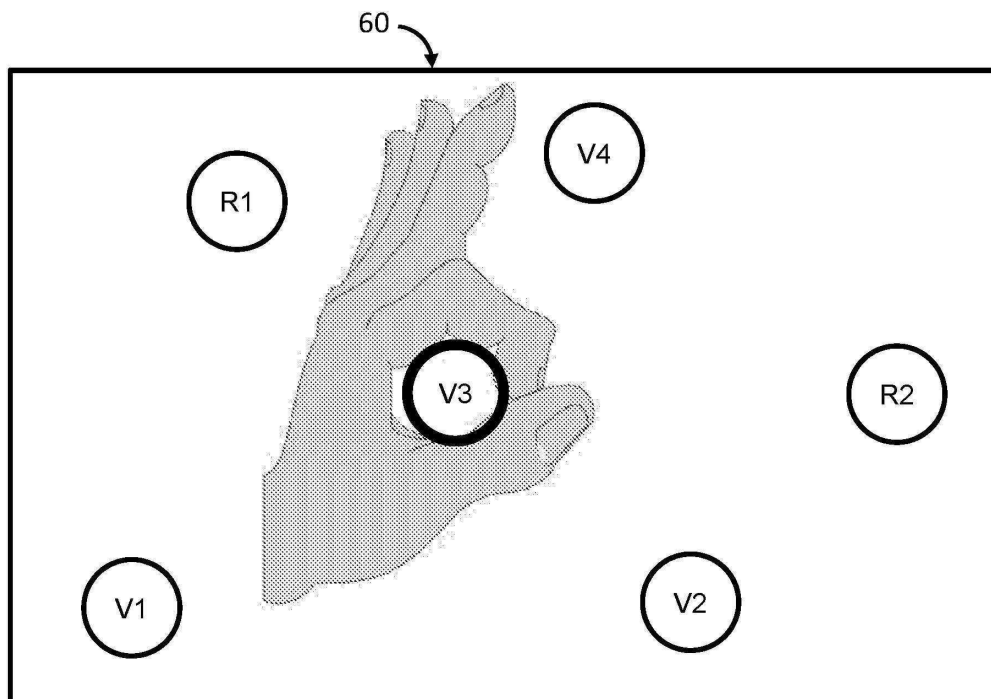
도면5



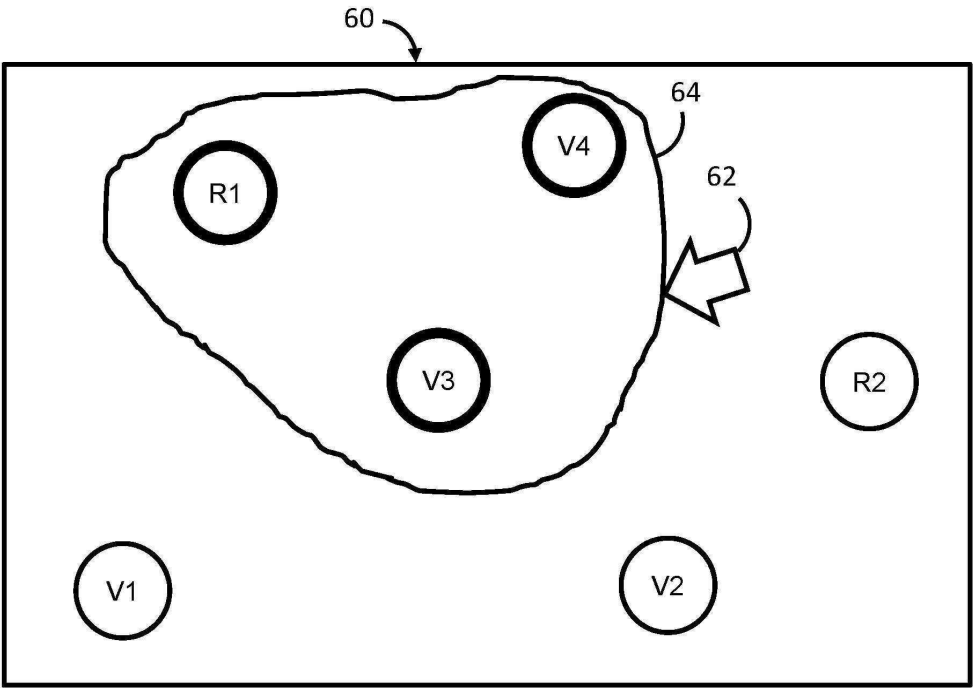
도면6



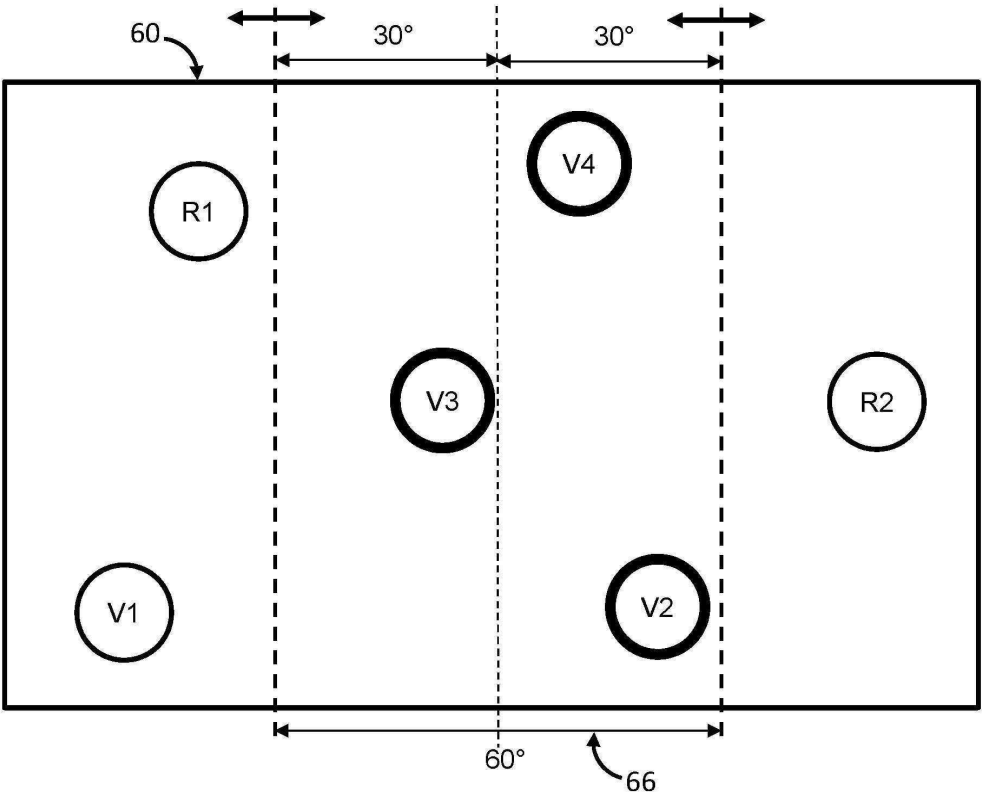
도면7



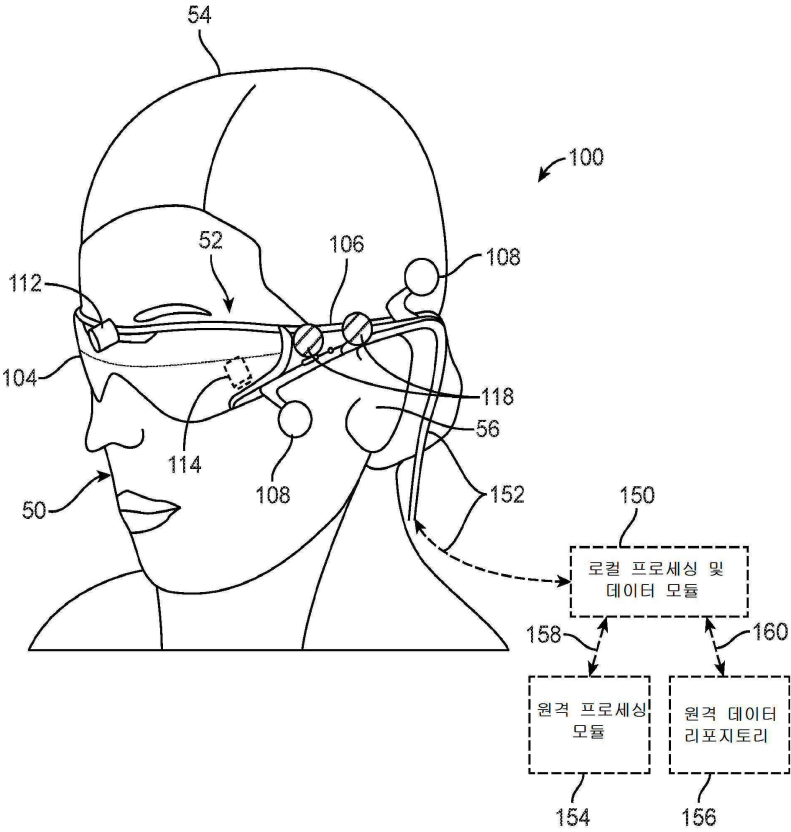
도면8



도면9

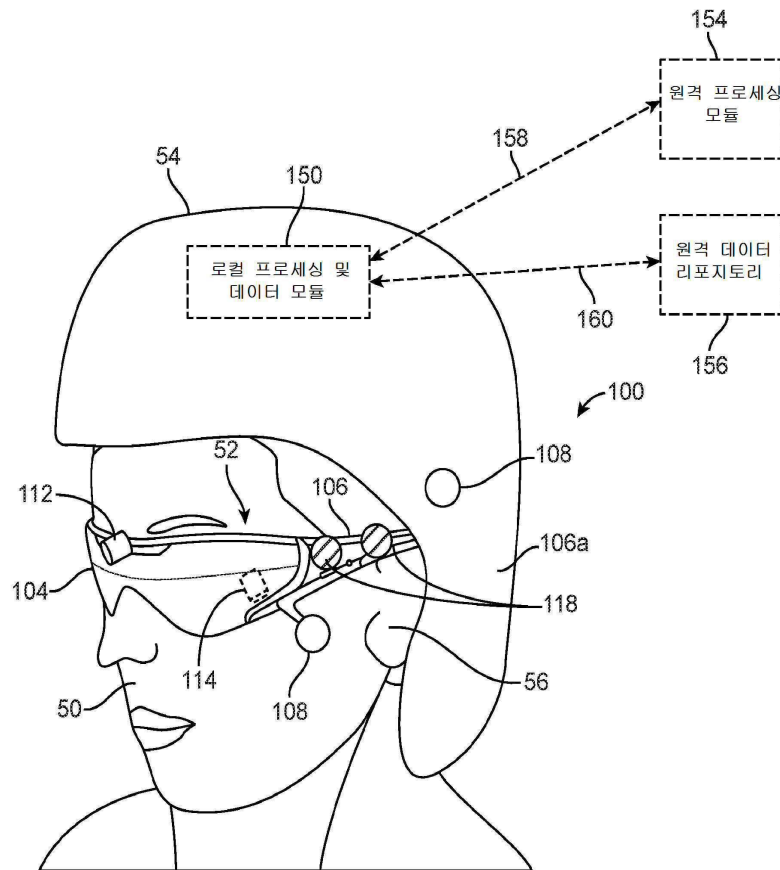


도면10a

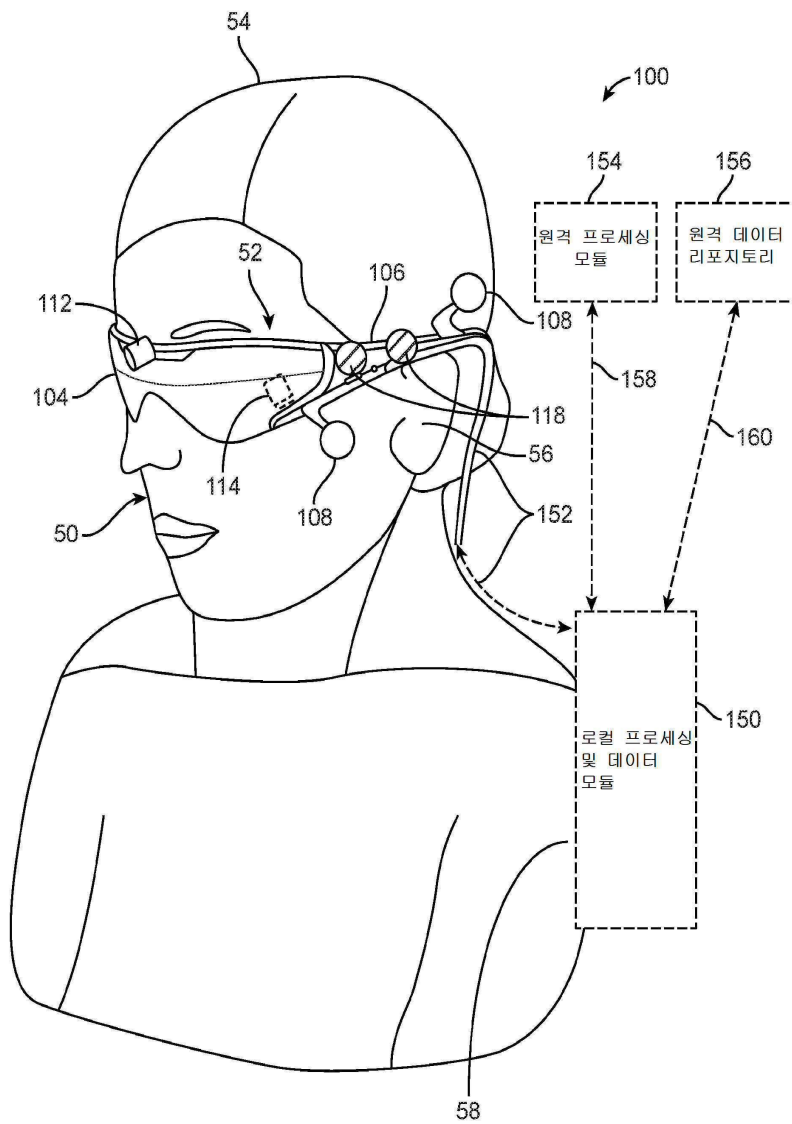




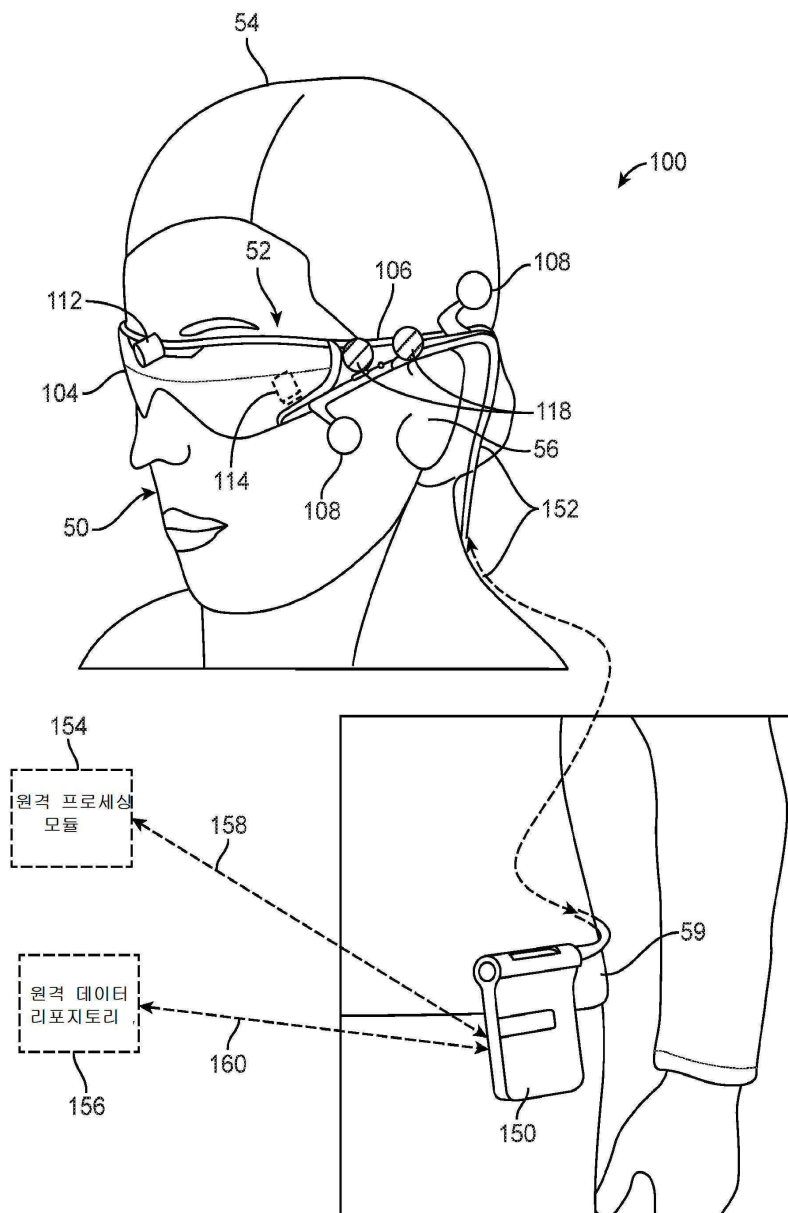
도면10b



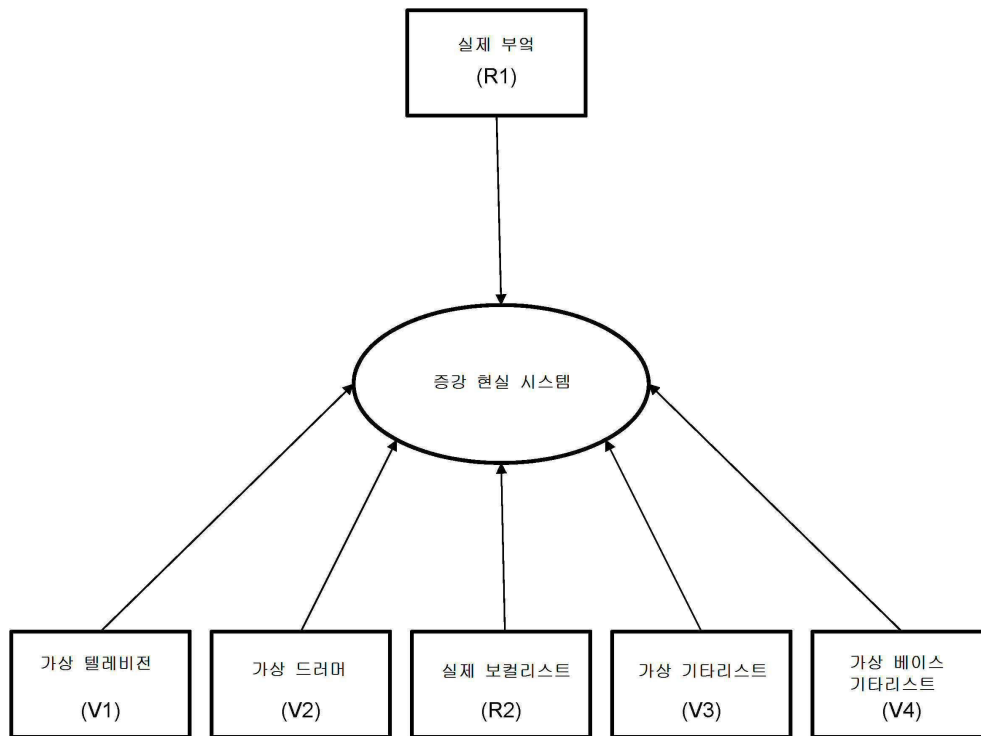
도면10c



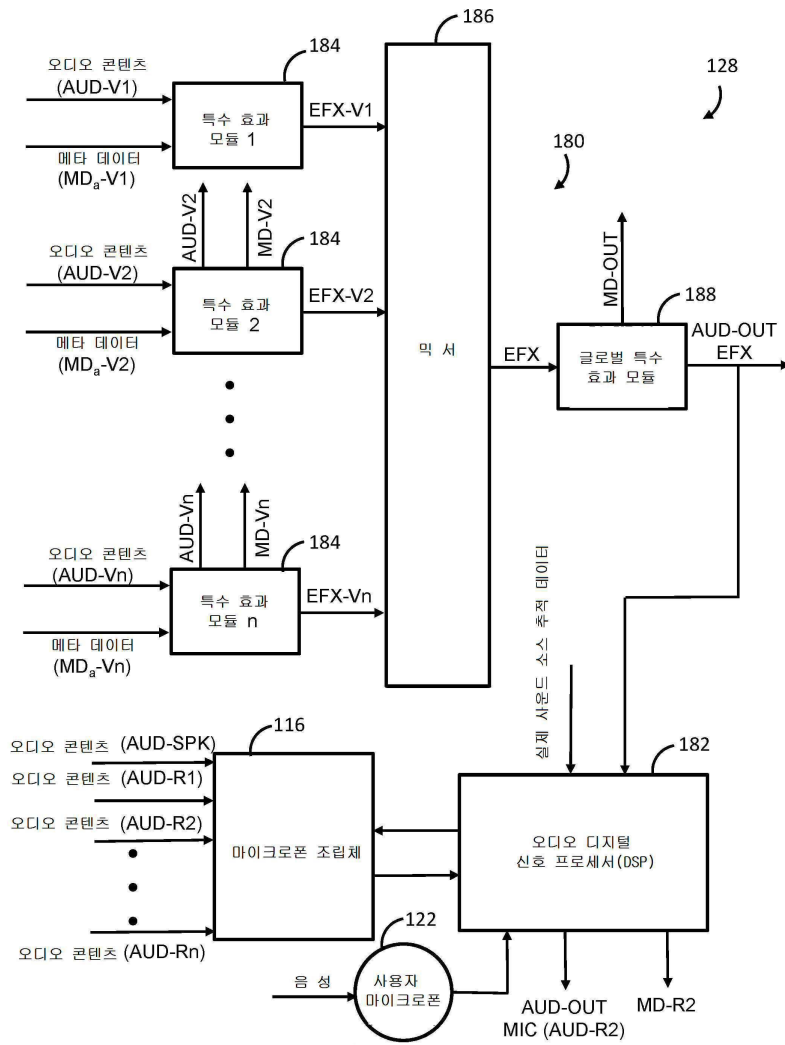
도면10d



도면11



도면12

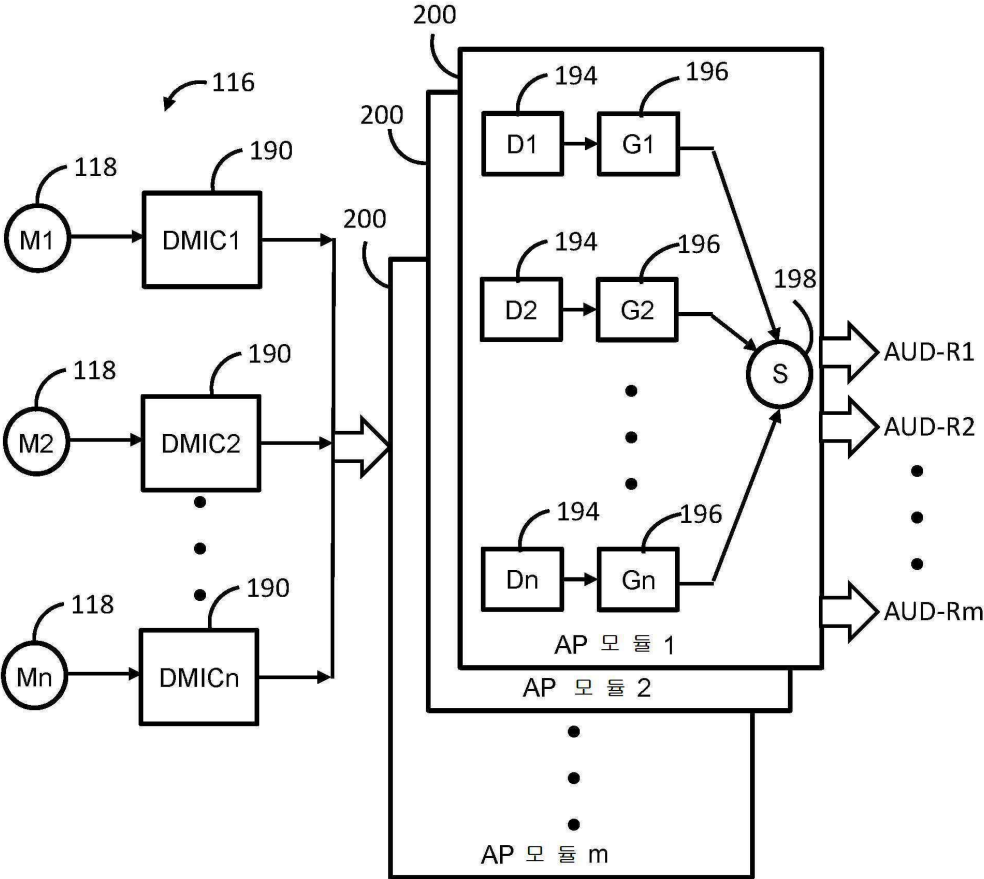


도면13

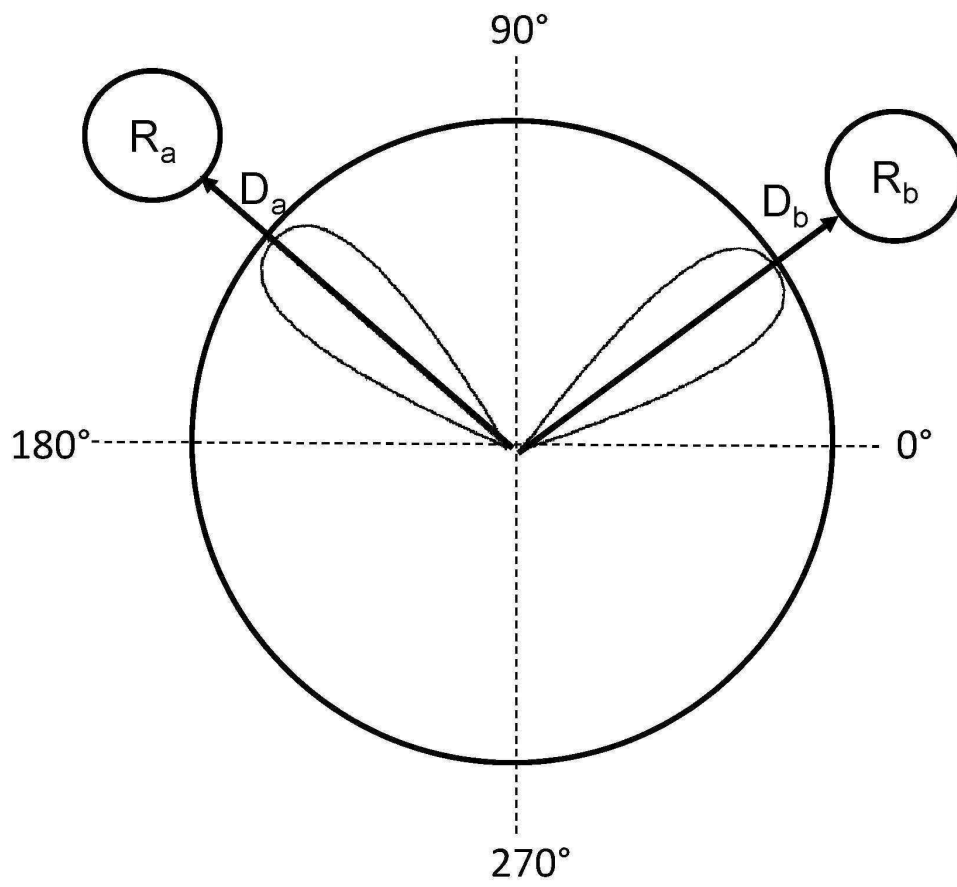
130

메모리		
메타-데이터	오디오 콘텐츠	비디오 콘텐츠
MD-OUT		
MD-V2	AUD-V2	VID-V2
MD-V3	AUD-V3	VID-V3
MD-V4	AUD-V4	VID-V4
MD-R2	AUD-MIC	VID-R2

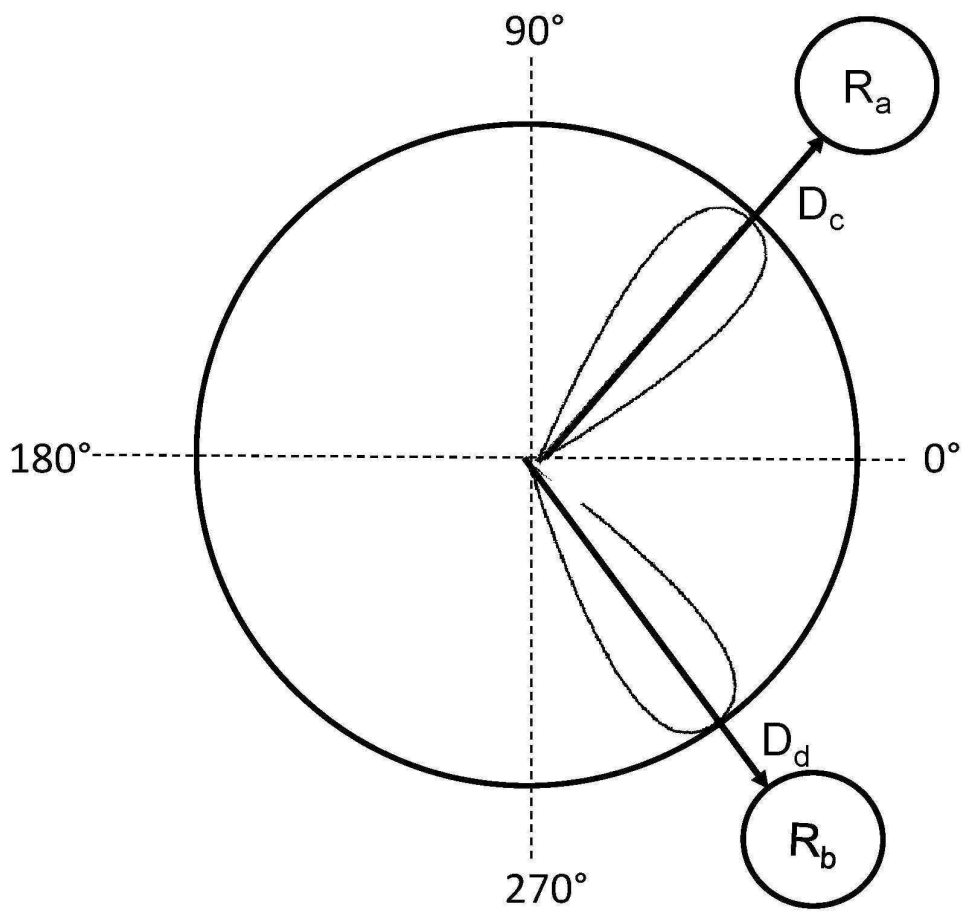
도면14



도면15a

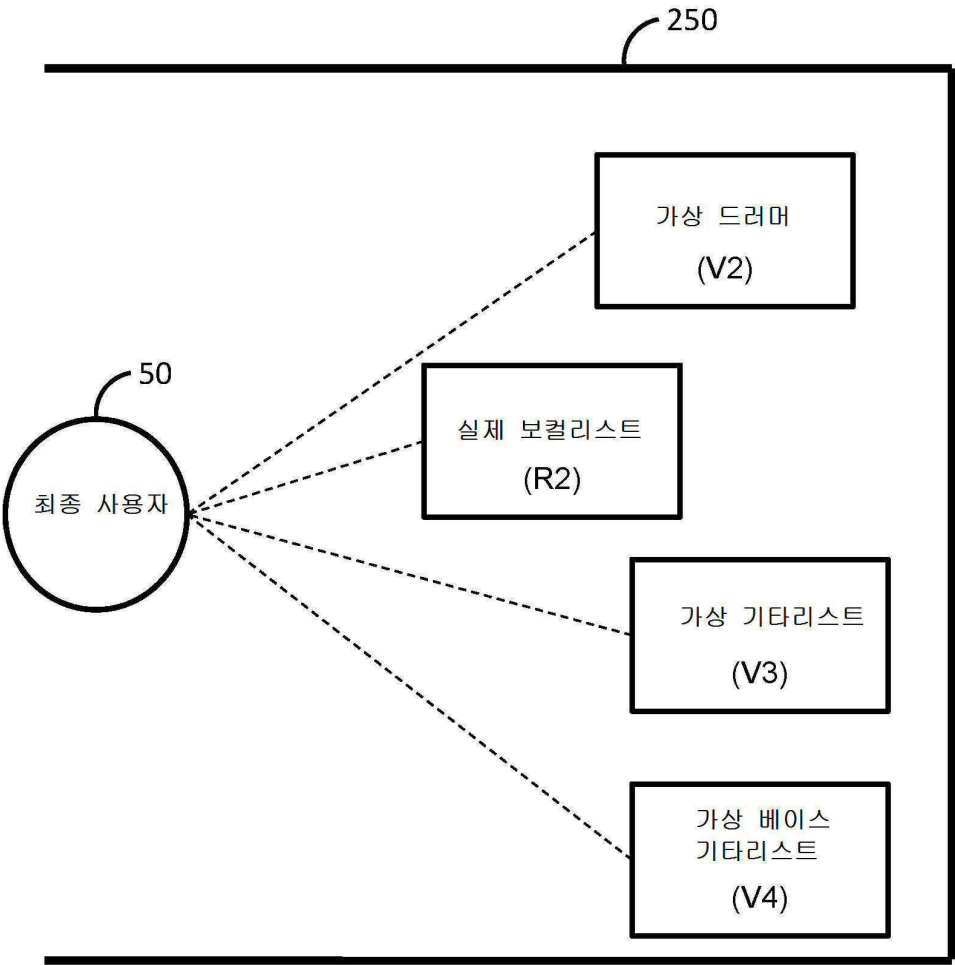


도면15b

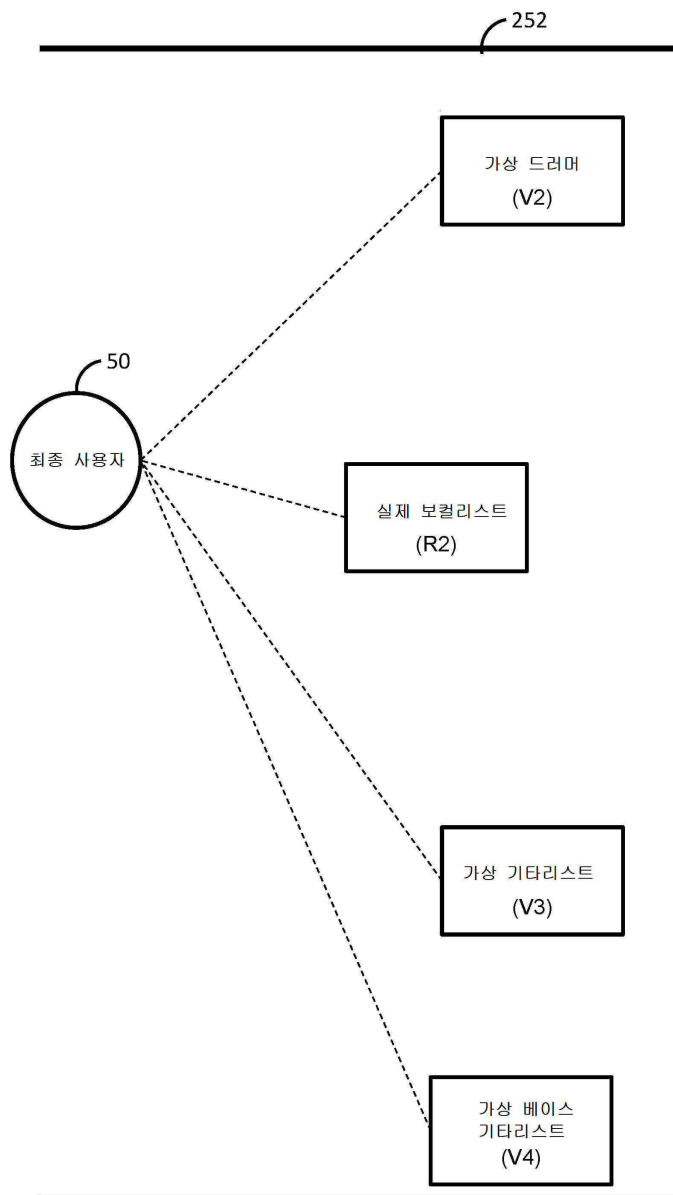




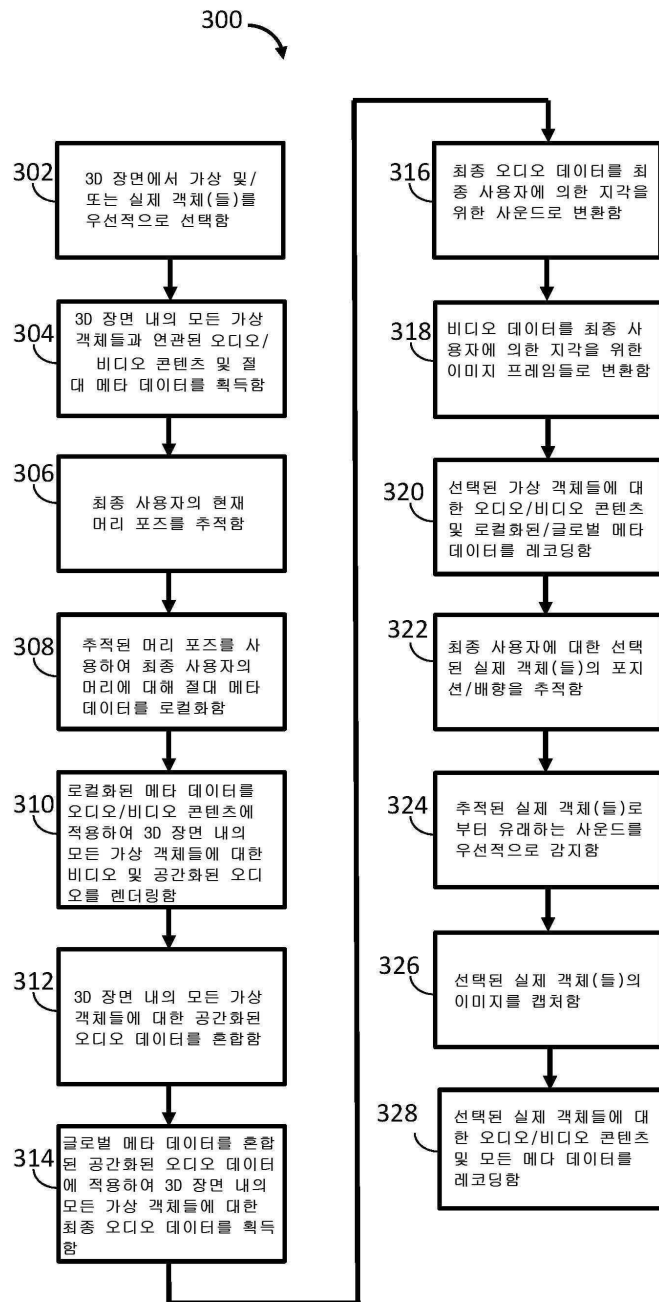
도면16a



도면16b



도면17



도면18

