



**(19) 대한민국특허청(KR)**  
**(12) 공개특허공보(A)**

(11) 공개번호 10-2012-0061869  
(43) 공개일자 2012년06월13일

(51) 국제특허분류(Int. Cl.)  
H04N 21/233 (2011.01)  
(21) 출원번호 10-2012-7006217  
(22) 출원일자(국제) 2010년08월13일  
심사청구일자 없음  
(85) 번역문제출일자 2012년03월08일  
(86) 국제출원번호 PCT/US2010/045530  
(87) 국제공개번호 WO 2011/020065  
국제공개일자 2011년02월17일  
(30) 우선권주장  
61/233,931 2009년08월14일 미국(US)

(71) 출원인  
에스알에스 랩스, 인크.  
미국 92705 캘리포니아주 산타 아나 데일러 스트리트 2909  
(72) 발명자  
트레이시 제임스  
미국 캘리포니아주 92705 산타 아나 다일러 스트리트 2909  
카치아노스 데미스  
미국 캘리포니아주 92346 하일랜드 처치 스트리트 6758  
크래머 앨런 디.  
미국 캘리포니아주 92680 투스틴 샤넬 드라이브 17661  
(74) 대리인  
신정건, 김태홍

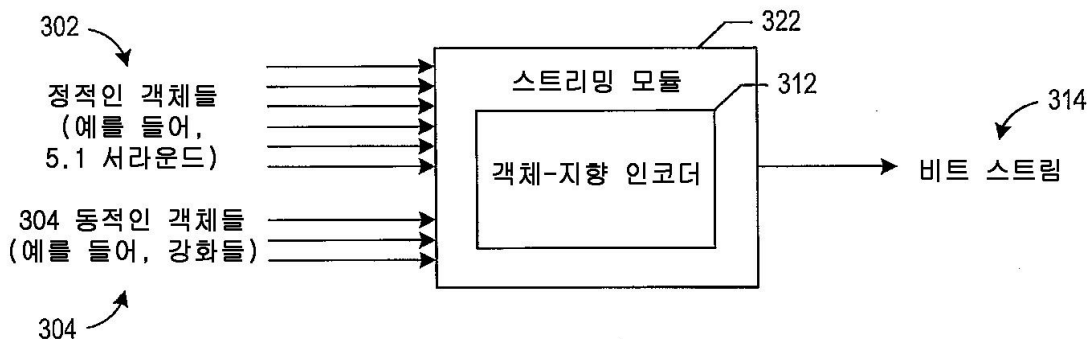
전체 청구항 수 : 총 17 항

(54) 발명의 명칭 **객체-지향 오디오 스트리밍 시스템**

**(57) 요약**

객체-지향 오디오를 제공하기 위한 시스템들 및 방법들이 기술된다. 오디오 객체들은 사운드 소스들을 위치, 속도, 배향성 등과 같은 그 오디오 소스들의 속성들과 연관시킴으로써 생성될 수 있다. 오디오 객체들은 사운드를 분배하기 위해 예를 들어, 네트워크를 통해 오디오 객체들을 클라이언트 디바이스에 스트리밍함으로써 채널에 더하여, 또는 그 대신에 이용될 수 있다. 객체들은 연관된 2 또는 3차원 좌표들을 이용하여 공간적으로 자신의 위치들을 정의할 수 있다. 객체들은 이용 가능한 네트워크 또는 클라이언트 디바이스 자원들에 기초하여 클라이언트 디바이스에 적응적으로 스트리밍될 수 있다. 클라이언트 디바이스 상의 렌더기는 객체들을 어떻게 렌더링할지를 결정하기 위해 객체들의 속성들을 이용할 수 있다. 렌더기는 또한 클라이언트 디바이스의 렌더링 환경에 관한 정보에 기초하여 객체들의 재생을 적응시킬 수 있다. 오디오 객체 생성 기법들의 다양한 예들이 또한 기술된다.

**대표도**



## 특허청구의 범위

### 청구항 1

객체-지향 오디오 스트림을 생성하는 방법에 있어서,

오디오 스트림 내에 포함시켜 전송하기 위해 오디오 객체를 선택하는 단계 - 상기 오디오 객체는 오디오 신호 데이터와 객체 메타데이터를 포함하고, 상기 객체 메타데이터는 하나 이상의 객체 속성을 포함함 -;

상기 객체 메타데이터를 포함하는 객체 헤더를 조립하는 단계;

상기 오디오 신호 데이터를 포함하는 오디오 페이로드를 제공하는 단계;

하나 이상의 프로세서를 사용해서 상기 오디오 스트림의 적어도 일부분을 형성하도록 상기 객체 헤더와 상기 오디오 페이로드를 조합하는 단계; 및

네트워크를 통해 상기 오디오 스트림을 전송하는 단계를

포함하는, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 2

제1항에 있어서, 상기 전송하는 단계는 상기 네트워크를 통해 단일 스트림으로서 상기 오디오 스트림을 전송하는 단계를 포함하는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 3

제1항에 있어서, 상기 하나 이상의 객체 속성은 상기 오디오 객체의 위치, 상기 오디오 객체의 속도, 상기 오디오 객체의 폐색(occlusion)과, 상기 오디오 객체와 연관된 환경 중 적어도 하나를 포함하는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 4

제1항에 있어서, 상기 조합하는 단계는 복수의 가변 길이 프레임으로부터 상기 오디오 스트림을 형성하는 단계를 포함하고, 각 프레임의 길이는 각 프레임과 연관된 상기 객체 메타데이터의 양(amount)에 적어도 부분적으로 종속되는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 5

제1항에 있어서, 상기 네트워크를 통해 상기 오디오 스트림을 전송하기 전에 상기 오디오 스트림을 압축하는 단계를 더 포함하는, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 6

제1항에 있어서, 상기 오디오 객체는 정적(static) 객체를 포함하는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 7

제6항에 있어서, 상기 정적 객체는 오디오 채널을 나타내는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 8

제6항에 있어서, 상기 오디오 스트림 내에 동적 오디오 객체를 배치시키는 단계를 더 포함하고, 상기 동적 오디오 객체는 상기 정적 객체를 강화시키도록(enhance) 구성된 강화 데이터를 포함하는 것인, 객체-지향 오디오 스트림을 생성하는 방법.

### 청구항 9

제1항에 있어서, 상기 오디오 스트림 내의 중복(redundant) 객체 메타데이터를 감소시키는 단계를 더 포함하는, 객체-지향 오디오 스트림을 생성하는 방법.

**청구항 10**

객체-지향 오디오 스트림을 생성하기 위한 시스템에 있어서,

하나 이상의 프로세서 내에 구현된 객체-지향 스트리밍 모듈을 포함하고, 상기 객체-지향 스트리밍 모듈은:

사운드 소스를 나타내는 오디오 객체를 선택하고 - 상기 오디오 객체는 오디오 신호 데이터와 객체 메타데이터를 포함하고, 상기 객체 메타데이터는 상기 사운드 소스의 하나 이상의 속성을 포함함 -;

단일 객체-지향 오디오 스트림의 적어도 일부분을 형성하도록 상기 오디오 신호 데이터와 함께 상기 객체 메타데이터를 인코딩하고;

네트워크를 통해 상기 객체-지향 오디오 스트림을 전송하도록 구성된, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 11**

제10항에 있어서, 상기 객체-지향 스트리밍 모듈은 제2 오디오 객체를 상기 객체-지향 오디오 스트림 내에 삽입하도록 더 구성되고, 상기 제2 오디오 객체는 오디오 페이로드를 포함하지 않고 제2 객체 메타데이터만을 포함하는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 12**

제11항에 있어서, 상기 제2 오디오 객체의 제2 객체 메타데이터는 환경 정의 데이터를 포함하는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 13**

제10항에 있어서, 상기 객체-지향 스트리밍 모듈은 상기 객체 메타데이터와 상기 오디오 신호 데이터 중 하나 또는 모두를 적어도 압축함으로써 상기 오디오 신호 데이터와 함께 상기 객체 메타데이터를 인코딩하도록 더 구성되는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 14**

제10항에 있어서, 상기 사운드 소스의 하나 이상의 속성은 상기 사운드 소스의 위치를 포함하는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 15**

제14항에 있어서, 상기 사운드 소스의 위치는 상기 오디오 객체와 연관된 비디오의 카메라 뷰(view)에 대해 결정되는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 16**

제10항에 있어서, 상기 사운드 소스의 하나 이상의 속성은

상기 오디오 객체에 의해 표현되는 상기 사운드 소스의 위치;

상기 사운드 소스의 속도;

상기 사운드 소스의 배향성(directionality);

상기 사운드 소스의 페색; 및

상기 사운드 소스와 연관된 환경

중 두 개 이상을 포함하는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**청구항 17**

제10항에 있어서, 상기 객체-지향 스트리밍 모듈은 상기 오디오 스트림 내의 중복 객체 메타데이터를 감소시키도록 더 구성되는 것인, 객체-지향 오디오 스트림을 생성하기 위한 시스템.

**명세서**

**기술분야**

[0001] 관련 출원

[0002] 이 출원은 미국의 35 U.S.C. § 119(e) 하에서 2009년 8월 14일 출원되고, 발명의 명칭이 "Production, Transmission, Storage and Rendering System for Multi-Dimensional Audio" 이며, 그 개시내용 전체가 여기에 참조로서 포함되는 미국 가특허 번호 제 61/233,931 호의 우선권의 이익을 청구한다.

**배경기술**

[0003] 스테레오 및 서라운드 사운드와 같은 기존의 오디오 분배 시스템들은 생산의 지점으로부터 재생 환경으로 고정된 수의 채널들을 구현하는 유연성 없는 패러다임을 기반으로 한다. 전체 오디오 체인에 걸쳐서, 기존에는 물리적으로 전송되거나 기록되는 채널들의 수와 생성되는 채널들의 수 사이에서 1 대 1 대응이 되어왔다. 몇몇 경우들에서, 이용 가능한 채널들의 수는 전송 스트림에서 제공되는 수보다 적은 재현 채널들을 갖는 재생 구성들을 수용하기 위해 믹스-다운(mix-down)으로서 알려진 프로세스를 통해 감소되었다. 믹스-다운의 공통적인 예들은 스테레오를 단일 스피커를 통한 재현을 위해 모노로 믹싱(mixing)하고 다중-채널 서라운드 사운드를 2개의 스피커 재생을 위해 스테레오로 믹싱한다.

**발명의 내용**

**해결하려는 과제**

[0004] 오디오 분배 시스템들은 또한 3D 비디오 애플리케이션들에 대해서 적합하지 않은데 그 이유는 이들은 3-차원 공간에서 정확하게 사운드를 랜더링할 수 없기 때문이다. 이 시스템들은 스피커들의 수 및 위치에 의해 그리고 음향심리학 원리(psychoacoustic principle)들이 일반적으로 무시된다는 사실에 의해 제한된다. 그 결과, 대부분의 정교한 사운드 시스템들은 진정한 3D 또는 다중차원 프리젠테이션에 근접하지 않는, 단지 음향 공간의 대충의 시뮬레이션(rough simulation)이다.

**과제의 해결 수단**

[0005] 객체-지향 오디오를 제공하기 위한 시스템들 및 방법들이 기술된다. 특정한 실시예들에서, 오디오 객체들은 사운드 소스들을 위치, 속도, 배향성 등과 같은 그 오디오 소스들의 속성들과 연관시킴으로써 생성된다. 오디오 객체들은 사운드를 분배하기 위해 예를 들어, 네트워크를 통해 오디오 객체들을 클라이언트 디바이스에 스트리밍함으로써 채널에 더하여, 또는 그 대신에 이용될 수 있다. 객체들은 연관된 2 또는 3차원 좌표들을 이용하여 공간적으로 자신의 위치들을 정의할 수 있다. 객체들은 이용 가능한 네트워크 또는 클라이언트 디바이스 자원들에 기초하여 클라이언트 디바이스에 적응적으로 스트리밍될 수 있다. 클라이언트 디바이스 상의 랜더기는 객체들을 어떻게 랜더링할지를 결정하기 위해 객체들의 속성들을 이용할 수 있다. 랜더기는 또한 클라이언트 디바이스의 랜더링 환경에 관한 정보에 기초하여 객체들의 재생을 적응시킬 수 있다. 오디오 객체 생성 기법들의 다양한 예들이 또한 기술된다.

[0006] 특정 실시예에서, 객체-지향 오디오 스트림을 생성하는 방법은 오디오 스트림 내에 포함시켜 전송하기 위해 오디오 객체를 선택하는 단계를 포함하고, 오디오 객체는 오디오 신호 데이터와 객체 메타데이터를 포함한다. 객체 메타데이터는 하나 이상의 객체 속성을 포함할 수 있다. 방법은 객체 메타데이터를 갖는 객체 헤더를 조립하는 단계와, 오디오 신호 데이터를 갖는 오디오 페이로드를 제공하는 단계와, 하나 이상의 프로세서를 사용하여 오디오 스트림의 적어도 일부분을 형성하도록 객체 헤더와 오디오 페이로드를 조합하는 단계와, 네트워크를 통해 오디오 스트림을 전송하는 단계를 더 포함할 수 있다.

[0007] 다양한 구현에서, 객체-지향 오디오 스트림을 생성하기 위한 시스템은 하나 이상의 프로세서 내에 구현된 객체-지향 스트리밍 모듈을 포함할 수 있다. 객체-지향 스트리밍 모듈은 사운드 소스를 나타내는 오디오 객체를 선택할 수 있고, 오디오 객체는 오디오 신호 데이터와 객체 메타데이터를 포함한다. 객체 메타데이터는 사운드 소스의 하나 이상의 속성을 포함할 수 있다. 객체-지향 스트리밍 모듈은 단일 객체-지향 오디오 스트림의 적어도 일부분을 형성하도록 오디오 신호 데이터와 함께 객체 메타데이터를 또한 인코딩할 수 있다. 또한, 객체-지향 스트리밍 모듈은 네트워크를 통해 객체-지향 오디오 스트림을 전송할 수 있다.

[0008] 본 개시를 요약하기 위해, 본 발명들의 특정한 양상들, 이점들 및 신규한 특징들이 여기서 기술되었다. 모든 이러한 이점들이 반드시 여기서 개시된 발명들의 임의의 특정한 실시예에 따라 달성될 수 있는 것은 아니라는 점이 이해될 것이다. 따라서, 여기서 개시된 발명들은 여기서 교시되거나 제안될 수 있는 바와 같은 다른 이점들을 반드시 달성하지 않고 여기서 교시된 바와 같은 하나의 이점 또는 일군의 이점들을 달성하거나 최적화하는 방식으로 구현되거나 실행될 수 있다.

[0009] 도면들 전체에 걸쳐서, 참조 번호들은 참조되는 엘리먼트들 간의 대응을 표시하는데 재사용된다. 도면들은 여기서 기술된 본 발명들의 실시예들을 예시하도록 제공되며 그 범위를 제한하도록 제공되지 않는다.

**발명의 효과**

[0010] 현재의 시스템들과 달리, 본 발명의 시스템(100A, 100B)은 유연하며 적응 가능하다.

**도면의 간단한 설명**

- [0011] 도 1a 및 도 1b는 객체-지향 오디오 시스템들의 실시예들을 예시하는 도면.
- 도 2는 객체-지향 오디오 시스템의 다른 실시예를 예시하는 도면.
- 도 3은 여기서 기술된 객체-지향 오디오 시스템들 중 임의의 시스템에서 이용하기 위한 스트리밍 모듈의 실시예를 예시하는 도면.
- 도 4는 객체-지향 오디오 스트리밍 포맷의 실시예를 예시하는 도면.
- 도 5a는 오디오 스트림 어셈블리 프로세스의 실시예를 예시하는 도면.
- 도 5b는 오디오 스트림 렌더링 프로세스의 실시예를 예시하는 도면.
- 도 6은 적응형 오디오 객체 스트리밍 시스템의 실시예를 예시하는 도면.
- 도 7은 적응형 오디오 객체 스트리밍 프로세스의 실시예를 예시하는 도면.
- 도 8은 적응형 오디오 객체 렌더링 프로세스의 실시예를 예시하는 도면.
- 도 9는 객체-지향 오디오 캡처를 위한 예시적인 장면을 예시하는 도면.
- 도 10은 객체-지향 오디오 캡처를 위한 시스템의 실시예를 예시하는 도면.
- 도 11은 객체-지향 오디오 캡처를 위한 프로세스의 실시예를 예시하는 도면.

**발명을 실시하기 위한 구체적인 내용**

[0012] I. 소개

[0013] 상술한 기존의 시스템들이 갖는 문제점들 외에, 오디오 분배 시스템들은 청취자의 재생 환경을 충분하게 고려하지 않는다. 대신에, 오디오 시스템들은 환경, 청취자 선호도 또는 음향 심리학적 원리들의 구현에 대한 어떠한 보상도 없이 특정된 수의 채널들을 최종 청취 환경에 전달하도록 설계된다. 이 기능들 및 성능들은 전통적으로 시스템 통합기에 맡겨진다.

[0014] 이 개시는 이러한 문제점들 중 적어도 일부를 해결하는 객체-지향 오디오를 스트리밍하기 위한 시스템들 및 방법들을 기술한다. 특정한 실시예들에서, 오디오 객체들은 위치, 속도, 배향성 등과 같이 사운드 소스들의 속성들을 그 사운드 소스들과 연관시킴으로써 생성된다. 오디오 객체들은 예를 들어, 네트워크를 통해 오디오 객체들을 클라이언트 디바이스에 스트리밍함으로써, 오디오를 분배하기 위한 채널들에 더하여 또는 그 대신에 이용될 수 있다. 특정한 실시예들에서, 이 객체들은 채널들 또는 채널들 간의 패닝된 위치들(panned positions)에 관련되지 않고, 오히려 2 또는 3차원 좌표들을 갖는 그들의 공간적 위치들을 정의한다. 클라이언트 디바이스 상의 렌더기는 객체들을 어떻게 렌더링할지를 결정하기 위해 객체들의 속성들을 이용할 수 있다.

[0015] 렌더기는 또한 이용 가능한 컴퓨팅 자원들에 기초하여 렌더링 및/또는 스트리밍을 적응시킴으로써 특정한 실시예들에서 렌더기의 환경을 고려할 수 있다. 유사하게, 오디오 객체들의 스트리밍은 이용 가능한 대역폭과 같은 네트워크 조건들에 기초하여 적응될 수 있다. 오디오 객체 생성 기법들의 다양한 예들이 또한 기술된다. 유리하게는, 여기서 기술되는 시스템들 및 방법들은 고속 오디오 채널 분배 모델과 연관된 단점들

을 감소시키거나 극복할 수 있다.

[0016] 개요로서, 도 1a 및 도 1b는 객체-지향 오디오 시스템들의 실시예들을 소개한다. 이후의 도면들은 이 객체-지향 오디오 시스템들에 의해 구현될 수 있는 기법들을 기술한다. 예를 들어, 도 2 내지 도 5b는 객체-지향 오디오를 스트리밍하기 위한 다양한 예시적인 기법들을 기술한다. 도 6 내지 도 8은 환경 및 네트워크 조건들에 기초하여 객체-지향 오디오를 적응적으로 스트리밍 및 랜더링하기 위한 예시적인 기법들을 기술한다. 도 9 내지 도 11은 예시적인 객체 생성 기법들을 기술한다.

[0017] 여기서 이용되는 바와 같이, 용어 "스트리밍(streaming)" 및 그 파생어들은 그들의 보편적인 의미를 갖는 것 외에, 하나의 컴퓨팅 시스템(예를 들어, 서버)으로부터 다른 컴퓨터 시스템(예를 들어, 클라이언트)으로의 콘텐츠의 분배를 의미할 수 있다. 용어 "스트리밍" 및 그 파생어들은 BitTorrent 및 관련 프로토콜들을 포함하는 임의의 다양한 프로토콜들을 이용하여 피어-투-피어 네트워크들을 통해 콘텐츠를 분배하는 것을 또한 지칭할 수 있다.

[0018] II. 객체-지향 오디오 시스템의 개요

[0019] 도 1a 및 도 1b는 객체-지향 오디오 시스템들(100A, 100B)의 실시예들을 예시한다. 객체-지향 오디오 시스템들(100A, 100B)은 컴퓨터 하드웨어 및/또는 소프트웨어로 구현될 수 있다. 유리하게는, 특정한 실시예들에서, 객체-지향 오디오 시스템들(100A, 100B)은 고정된 채널 모델로 속박됨 없이, 콘텐츠 생성기들이 오디오 객체들을 생성하고, 이러한 객체들을 스트리밍하고 객체들을 랜더링하는 것을 가능하게 할 수 있다.

[0020] 도 1a를 구체적으로 참조하면, 객체-지향 오디오 시스템(100A)은 오디오 객체 생성 시스템(110A), 콘텐츠 서버(120A)로 구현된 스트리밍 모듈(122A), 및 사용자 시스템(140)으로 구현된 랜더기(142A)를 포함한다. 오디오 객체 생성 시스템(110A)은 사용자가 오디오 객체들을 생성하고 수정하기 위한 기능을 제공할 수 있다. 콘텐츠 서버(120A) 상에 설치된 것으로 도시되는 스트리밍 모듈(122A)은 네트워크(130)를 통해 오디오 객체들을 사용자 시스템(140)으로 스트리밍하는데 이용될 수 있다. 네트워크(130)는 LAN, WAN, 인터넷 또는 이들의 조합들을 포함할 수 있다. 사용자 시스템(140) 상의 랜더기(142A)는 출력용 오디오 객체들을 하나 이상의 로스피커들로 랜더링할 수 있다.

[0021] 도시된 실시예들에서, 오디오 객체 생성 시스템(110A)은 객체 생성 모듈(114) 및 객체-지향 인코더(112A)를 포함한다. 객체 생성 모듈(114)은 예를 들어, 오디오 데이터를 오디오 데이터의 속성들과 연관시킴으로써 객체들을 생성하기 위한 기능을 제공할 수 있다. 임의의 타입의 오디오가 오디오 객체를 생성하는데 이용될 수 있다. 객체들로 생성되고 스트리밍될 수 있는 몇몇의 예들은 영화들, 텔레비전, 영화 트레일러(movie trailer), 음악, 음악 비디오들, 다른 온라인 비디오들, 비디오 게임들 등과 연관된 오디오를 포함할 수 있다.

[0022] 먼저, 오디오 데이터가 기록되거나 다른 방식으로 획득될 수 있다. 객체 생성 모듈(114)은 사용자가 오디오 데이터를 액세스, 편집, 또는 다른 방식으로 조작하는 것을 가능하게 하는 사용자 인터페이스를 제공할 수 있다. 오디오 데이터는 사운드 소스 또는 사운드 소스들의 컬렉션(collection)을 나타낼 수 있다. 사운드 소스들의 몇몇 예들은 다이얼로그(dialog), 배경 음악, 및 임의의 아이템(예를 들어, 차, 비행기, 또는 임의의 프롭(prop))에 의해 생성되는 사운드들을 포함한다. 보다 일반적으로, 사운드 소스는 임의의 오디오 클립(clip)일 수 있다.

[0023] 사운드 소스들은 객체 생성 모듈(114)이 객체를 생성하기 위해 오디오 데이터와 연관시킬 수 있는 하나 이상의 속성들을 가질 수 있다. 속성들의 예들은 사운드 소스의 위치, 사운드 소스의 속도, 사운드 소스의 배향성 등을 포함한다. 오디오 데이터가 기록된 시간을 반영하는 시간 속성과 같은 몇몇의 속성들은 오디오 데이터로부터 직접 획득될 수 있다. 다른 속성들은 오디오를 생성한 오디오 소스의 타입(예를 들어, 자동차 대 배우)과 같이 사용자에게 의해 객체 생성 모듈(114)에 공급될 수 있다. 또 다른 속성들은 다른 디바이스들로부터 객체 생성 모듈(114)에 의해 자동으로 입수될 수 있다. 예로서, 사운드 소스의 위치는 GPS(Global Positioning System) 디바이스 등으로부터 검색될 수 있거나 객체 생성 모듈(114)내로 입수될 수 있다. 속성들을 식별하기 위한 기법들 및 속성들의 부가적인 예들은 더 상세히 후술된다. 객체 생성 모듈(114)은 데이터베이스 또는 다른 데이터 스토리지를 포함할 수 있는 객체 데이터 저장소에 오디오 객체들을 저장할 수 있다.

- [0024] 객체-지향 인코더(112A)는 하나 이상의 오디오 객체들을 네트워크를 통한 전송에 적합한 오디오 스트림으로 인코딩할 수 있다. 일 실시예에서, 객체-지향 인코더(112A)는 연관된 속성 메타데이터와 함께 압축되지 않은 PCM(pulse code modulated) 오디오로서 오디오 객체들을 인코딩한다. 다른 실시예에서, 객체-지향 인코더(112A)는 또한 스트림을 생성할 때 압축을 객체들에 적용할 수 있다.
- [0025] 유리하게는, 특정한 실시예들에서, 객체-지향 인코더에 의해 생성된 오디오 스트림은 메타데이터 헤더 및 오디오 페이로드에 의해 표현되는 적어도 하나의 객체를 포함할 수 있다. 오디오 스트림은 객체 메타데이터 헤더들 및 오디오 페이로드들을 각각 포함할 수 있는 프레임들로 구성될 수 있다. 몇몇 객체들은 메타데이터만을 포함하고 어떠한 오디오 페이로드도 포함하지 않을 수 있다. 다른 객체들은 오디오 페이로드를 포함하지만 메타데이터를 거의 또는 전혀 포함하지 않을 수 있다. 이러한 객체들의 예들은 상세히 후술된다.
- [0026] 오디오 객체 생성 시스템(110A)은 인코딩된 오디오 객체들을 네트워크(도시되지 않음)를 통해 콘텐츠 서버(120A)에 공급할 수 있다. 콘텐츠 서버(120A)는 추후의 전송을 위해 인코딩된 오디오 객체들을 호스팅(host)할 수 있다. 콘텐츠 서버(120A)는 물리적인 컴퓨팅 디바이스들과 같은 하나 이상의 기계들을 포함할 수 있다. 콘텐츠 서버(120A)는 네트워크(130)를 통해 사용자 시스템들에 액세스 가능할 수 있다. 예를 들어, 콘텐츠 서버(120A)는 웹 서버, CDN(content delivery network)의 에지 노드 등일 수 있다.
- [0027] 사용자 시스템(140)은 오디오 콘텐츠를 요청하기 위해 콘텐츠 서버(120A)에 액세스할 수 있다. 이러한 요청의 수신에 응답하여, 콘텐츠 서버(120A)는 오디오 콘텐츠를 사용자 시스템(140)에 스트리밍, 업로딩, 또는 다른 방식으로 전송할 수 있다. 임의의 형태의 컴퓨팅 디바이스는 오디오 콘텐츠에 액세스할 수 있다. 예를 들어, 사용자 시스템(140)은 데스크톱, 랩톱, 태블릿, PDA(personal digital assistant), 텔레비전, 무선 핸드헬드 디바이스(예를 들어, 전화) 등일 수 있다.
- [0028] 사용자 시스템(140) 상의 랜더기(142A)는 인코딩된 오디오 객체들을 디코딩하여 출력용 오디오 객체들을 하나 이상의 로드스피커들로 랜더링할 수 있다. 랜더기(142A)는 오디오 객체들을 랜더링하기 위해 다양한 상이한 랜더링 특징들, 오디오 강화들, 음향 심리학적 강화들 등을 포함할 수 있다. 랜더기(142A)는 오디오 객체들을 어떻게 랜더링할지에 관한 단서들로서 오디오 객체들의 객체 속성들을 이용할 수 있다.
- [0029] 도 1b를 참조하면, 객체-지향 오디오 시스템(100B)은 오디오 객체 생성 시스템(110B), 콘텐츠 서버(120B) 및 사용자 시스템(140)과 같이 시스템(100A)의 다수의 특징들을 포함할 수 있다. 도시된 컴포넌트들의 기능은 상술한 것과 동일할 수 있으며, 특정한 차이들이 여기서 언급된다. 예를 들어, 도시된 실시예에서, 콘텐츠 서버(120B)는 사용자 시스템(140)에 스트리밍되는 객체 데이터의 양을 동적으로 적응시킬 수 있는 적응형 스트리밍 모듈(122B)을 포함한다. 유사하게, 사용자 시스템(140)은 오디오 스트리밍 및/또는 객체들이 사용자 시스템(140)에 의해 랜더링되는 방법을 적응시킬 수 있는 적응형 랜더기(142B)를 포함한다.
- [0030] 도 1b로부터 알 수 있는 바와 같이, 객체-지향 인코더(112B)는 오디오 객체 생성 시스템(110B)으로부터 콘텐츠 서버(120B)로 이동되었다. 도시된 실시예에서, 오디오 스트림들 대신에 오디오 객체들을 콘텐츠 서버(120B)에 업로드한다. 콘텐츠 서버(120B) 상의 적응형 스트리밍 모듈(122B)은 객체-지향 인코더(112B)를 포함한다. 그러므로 오디오 객체들의 인코딩은 도시된 실시예에서 콘텐츠 서버(120B) 상에서 수행된다. 대안적으로 오디오 객체 생성 시스템(110B)은 인코딩된 객체들을, 추가의 조작 및 추후의 재-인코딩을 위해 오디오 객체들을 디코딩하는 적응형 스트리밍 모듈(122B)로 스트리밍할 수 있다.
- [0031] 콘텐츠 서버(120B) 상에서 객체들을 인코딩함으로써, 적응형 스트리밍 모듈(122B)은 스트리밍 이전에, 객체들이 인코딩되는 방법을 동적으로 적응시킬 수 있다. 적응형 스트리밍 모듈(122B)은 네트워크 대역폭, 레이턴시 등과 같이 이용 가능한 네트워크(130) 자원들을 모니터링할 수 있다. 이용 가능한 네트워크 자원들에 기초하여, 적응형 스트리밍 모듈(122B)은 더 많은 또는 더 적은 오디오 객체들을 오디오 스트림으로 인코딩할 수 있다. 예를 들어, 네트워크 자원들이 더 많이 이용 가능하게 됨에 따라 적응형 스트리밍 모듈(122B)은 상대적으로 더 많은 오디오 객체들을 오디오 스트림으로 인코딩할 수 있고 그 반대도 가능하다.
- [0032] 적응형 스트리밍 모듈(122B)은 또한 그 개수가 아니라(또는 그 외에), 오디오 스트림으로 인코딩되는 객체들의 타입들을 조정할 수 있다. 예를 들어, 적응형 스트리밍 모듈(122B)은 네트워크 자원들이 제한될 때 더 낮은 우선순위 객체들(예를 들어, 특정한 배경 사운드들)이 아니라 더 높은 우선순위 객체들(예를 들어, 다이얼로그)을 인코딩할 수 있다. 객체 우선순위에 기초하여 스트리밍을 적응시키는 개념은 더 상세히 후술된다.
- [0033] 적응형 랜더기(142B)는 또한 오디오 객체들이 사용자 시스템(140)에 스트리밍되는 방법에 영향을 미칠 수 있다. 예를 들어, 적응형 랜더기(142B)는 사용자 시스템(140)에 스트리밍되는 오디오 객체들의 타입 및/또는

양을 제어하기 위해 적응형 스트리밍 모듈(122B)과 통신할 수 있다. 적응형 렌더기(142B)는 또한 재생 환경에 기초하여 오디오 스트림들이 렌더링되는 방법을 조정할 수 있다. 예를 들어, 대형 극장은 수십 개 또는 수백 개의 증폭기들 및 스피커들의 위치 및 성능들을 특정할 수 있는 반면에, 독립형 TV는 단지 2개의 증폭기 채널들 및 스피커들만이 이용 가능하다고 특정할 수 있다. 이런 정보에 기초하여, 시스템(100A, 100B)은 음향 필드 프리젠테이션을 최적화할 수 있다. 인입하는 오디오 스트림이 서술적(descriptive)일 수 있고 재생 환경의 물리적 특징들에 의존하지 않을 수 있기 때문에 시스템들(100A, 100B)의 다수의 상이한 타입들의 렌더링 특징들은 재현 자원들(reproducing resources) 및 환경에 의존하여 적용될 수 있다. 적응형 렌더기(142B)의 위의 및 다른 특징들은 더 상세히 후술된다.

[0034] 몇몇 실시예들에서, 여기서 기술되는 적응형 특징들은, 객체-지향 인코더(예를 들어, 인코더(112A))가 인코딩된 스트림을 적응형 스트리밍 모듈(122B)로 송신하는 경우조차도 구현될 수 있다. 새로운 오디오 스트림을 공중에서 어셈블리하는 대신에, 적응형 스트리밍 모듈(122B)은 컴퓨팅 자원들 또는 네트워크 자원들이 더 적게 이용 가능하게 될 때 오디오 스트림으로부터 객체를 제거하거나, 또는 다른 방식으로 오디오 스트림을 필터링할 수 있다. 예를 들어, 적응형 스트리밍 모듈(122B)은 렌더링하는데 상대적으로 덜 중요한 객체들에 대응하는 스트림으로부터 패킷들을 제거할 수 있다. 스트리밍 및/또는 렌더링에 대한 중요도를 객체들에 지정하기 위한 기법들은 더 상세히 후술된다.

[0035] 위의 실시예들로부터 알 수 있는 바와 같이, 오디오 분배 및 재생을 위한 개시된 시스템들(100A, 100B)은 오디오 콘텐츠의 초기 생성으로부터 청취자(들)의 지각 시스템으로의 전체 체인을 포함할 수 있다. 시스템들(100A, 100B)은 다중-차원 렌더링 시스템 또는 전송/저장에 있어서의 개념적인 개선들이 쉽게 통합될 수 있다는 점에서 스케일러블(scalable)하며 미래에도 경쟁력을 갖추게 될 수 있다. 시스템들(100A, 100B)은 또한 대형 포맷의 극장 기반 프리젠테이션으로부터 홈 시어러 구성들 및 독립형 TV 오디오 시스템들로 쉽게 스케일링될 수 있다.

[0036] 기존의 물리적인 채널 기반 시스템들과 대조적으로, 시스템들(100A, 100B)은 장면 내의 개별적인 컴포넌트들은 물론, 장면의 구조에 관한 정보를 제공하는 일련의 오디오 객체들로 오디오 콘텐츠의 제작물을 축약할 수 있다. 각 객체와 연관된 정보는 이용 가능한 자원들이 주어지면, 제공된 정보의 가장 정확한 프리젠테이션을 생성하기 위해 시스템들(100A, 100B)에 의해 이용될 수 있다. 이 자원들은 시스템들(100A, 100B)로의 부가적인 입력으로서 특정될 수 있다.

[0037] 물리적인 스피커들 및 증폭기들을 이용하는 것 외에, 시스템들(100A, 100B)은 음향 환경에서 청취자 몰입도를 강화하는 것을 물론, 비주얼(visual) 필드에서 그들의 위치에 정확하게 대응하는 3D 객체들의 포지셔닝을 구현하기 위해 음향 심리학적 프로세싱을 또한 통합시킬 수 있다. 이 프로세싱은 전송 스트림에 포함된 오디오 객체 정보의 프리젠테이션을 강화하거나, 또는 다른 방식으로 최적화하는데 이용 가능한 자원으로 시스템들(100A, 100B)(예를 들어, 렌더기(142))에 대해 또한 정의될 수 있다.

[0038] 스트림은 부가적인 정보가 임의의 시간에 부가될 수 있도록 확장 가능하게 되도록 설계된다. 렌더기(142A, 142B)는 일반적이거나 특정한 환경 및 자원 믹스(mix)를 지원하도록 설계될 수 있다. 오디오 재현에 있어서의 미래의 개선들 및 새로운 개념들은 자유로이 통합될 수 있으며 전송/저장 스트림에 포함된 동일한 서술적인 정보는 잠재적으로 더 정확한 렌더링에 활용된다. 시스템(100A, 100B)은 이전의 콘텐츠 및 렌더링 시스템과의 호환성을 유지하면서 임의의 미래의 물리적 또는 개념적인 개선들이 시스템(100A, 100B) 내의 임의의 지점에 쉽게 통합될 수 있다는 정도로 축약된다. 현재의 시스템들과 달리, 시스템(100A, 100B)은 유연하며 적용 가능하다.

[0039] 설명의 편의를 위해, 이 명세서는 주로 네트워크를 통한 오디오의 스트리밍의 문맥에서 객체-지향 오디오 기법들을 기술한다. 그러나 객체-지향 오디오 기법들은 또한 비-네트워크 환경들에서 구현될 수 있다. 예를 들어, 객체-지향 오디오 스트림은 DVD 디스크, 블루-레이 디스크 등과 같은 컴퓨터-판독 가능한 저장 매체에 상에 저장될 수 있다. 미디어 재생기(예를 들어, 블루-레이 재생기)는 디스크 상에 저장된 객체-지향 오디오 스트림을 재생할 수 있다. 객체-지향 오디오 패키지는 또한 사용자 시스템 상의 로컬 스토리지로 다운로드되고 그 후 로컬 스토리지로부터 재생될 수 있다. 다수의 다른 변형들이 가능하다.

[0040] 도 1a 및 도 1b에 대해 기술된 특정한 컴포넌트들의 기능은 조합되거나, 변형되거나, 또는 생략될 수 있다는 것이 인지되어야 한다. 예를 들어, 일 구현에서, 오디오 객체 생성 시스템(110)은 콘텐츠 서버(120) 상에 구현될 수 있다. 오디오 스트림들은 오디오 객체 생성 시스템(110)으로부터 사용자 시스템(140)으로 직접 스트리밍될 수 있다. 다수의 다른 구성들이 가능하다.



[0041] III. 오디오 객체 스트리밍 실시예들

[0042] 오디오 객체 스트림들의 더욱 상세한 실시예들이 도 2 내지 도 5b에 대하여 이제 기술될 것이다. 도 2를 참조하면, 객체-지향 오디오 시스템(200)의 다른 실시예가 도시된다. 시스템(200)은 상술된 시스템들(100A, 100B)의 특징들 중 임의의 특징을 구현할 수 있다. 시스템(200)은 디코딩되고, 렌더링되고 하나 이상의 스피커들에 의해 출력될 수 있는 객체-지향 오디오 스트림을 생성할 수 있다.

[0043] 시스템(200)에서, 오디오 객체들(202)은 객체-지향 인코더(212)에 제공된다. 객체-지향 인코더(212)는 상술된 바와 같이 콘텐츠 서버상의 오디오 콘텐츠 생성 시스템 또는 스트리밍 모듈에 의해 구현될 수 있다. 객체-지향 인코더(212)는 오디오 객체들을 비트 스트림(214)으로 인코딩하고 및/또는 압축할 수 있다. 객체-지향 인코더(212)는 (예를 들어, MP3 파일들을 생성하기 위해) MPEG(Moving Picture Experts Group) 표준들 중 임의의 표준에 기초한 압축 기법들을 포함하는 임의의 코덱 또는 압축 기법을 이용하여 객체들을 인코딩할 수 있다.

[0044] 특정한 실시예들에서, 객체-지향 인코더(212)는 상이한 오디오 객체들에 대해 메타데이터 헤더들 및 오디오 페이로드들을 갖는 단일 비트 스트림(214)을 생성한다. 객체-지향 인코더(212)는 네트워크를 통해 비트 스트림(214)을 전송할 수 있다(예를 들어, 도 1b를 참조한다). 사용자 시스템상에 구현된 디코더(220)는 비트 스트림(214)을 수신할 수 있다. 디코더(220)는 비트 스트림(214)을 그 성분 오디오 객체들(202)로 디코딩할 수 있다. 디코더(220)는 오디오 객체들(202)을 렌더기(242)에 제공한다. 몇몇 구현들에서, 렌더기(242)는 디코더(220)의 기능을 직접 구현할 수 있다.

[0045] 렌더기(242)는 오디오 객체들을 하나 이상의 스피커들(250) 상에서의 재생에 적합한 오디오 신호들(244)로 렌더링할 수 있다. 상술된 바와 같이, 렌더기(142A)는 오디오 객체들을 어떻게 렌더링할지에 관한 단서들로서 오디오 객체들의 객체 속성들을 이용할 수 있다. 유리하게는, 특정한 실시예들에서, 오디오 객체들은 이러한 속성들을 포함하기 때문에, 렌더기(142A)의 기능은 오디오 객체들의 포맷을 변경하지 않고 변경될 수 있다. 예를 들어, 일 타입의 렌더기(142A)는 하나의 스피커로부터 다른 스피커로 오디오를 패닝하기 위해 오디오 객체의 위치 속성을 이용할 수 있다. 제 2 렌더기(142A)는 음향 심리학적 강화가 렌더기(142A)에 대해 이용 가능하다는 결정에 응답하여 오디오 객체들에 대해 3D 음향 심리학적 필터링을 수행하기 위해 동일한 위치 속성을 이용할 수 있다. 일반적으로 렌더기(142A)는 최상의 가능한 프리젠테이션을 생성하는데 이용 가능한 자원들 모두 또는 그 일부를 고려할 수 있다. 렌더링 기술이 개선됨에 따라, 오디오 객체들의 선제하는 포맷(preexisting format)의 이점을 취하는 부가적인 렌더기들(142A) 또는 렌더링 자원들이 사용자 시스템(140)에 부가될 수 있다.

[0046] 상술된 바와 같이, 객체-지향 인코더(212) 및/또는 렌더기(242)는 또한 적응형 특징들을 가질 수 있다.

[0047] 도 3은 여기서 기술된 객체-지향 오디오 시스템들 중 임의의 시스템과 함께 사용하기 위한 스트리밍 모듈(322)의 실시예를 예시한다. 스트리밍 모듈(322)은 객체-지향 인코더(312)를 포함한다. 스트리밍 모듈(322) 및 인코더(312)는 하드웨어로 및/또는 소프트웨어로 구현될 수 있다. 도시된 실시예는 상이한 타입들의 오디오 객체들이 단일 비트 스트림(314)으로 어떻게 인코딩되는지를 예시한다.

[0048] 도시된 예시적인 스트리밍 모듈(322)은 2개의 상이한 타입들의 객체들? 정적 객체들(302) 및 동적 객체들(304)을 수신할 수 있다. 정적 객체들(302)은 5.1 채널 서라운드 사운드와 같은 오디오의 채널들을 나타낼 수 있다. 각 채널은 정적 객체(302)로서 표현될 수 있다. 몇몇의 콘텐츠 생성기들은 시스템들(100A, 100B)의 객체-기반 기능에 더하여 또는 그 대신에 채널들을 이용하고자 할 수 있다. 정적 객체들(302)은 이 콘텐츠 생성기들이 채널들을 이용하는 방법을 제공하여 기존의 고정된 채널 시스템들과의 역방향 호환성을 용이하게 하고 채택의 용이성(ease of adoption)을 조장한다.

[0049] 동적인 객체들(304)은 정적인 객체들(302) 외에 또는 그 대신에 이용될 수 있는 임의의 객체들을 포함할 수 있다. 동적인 객체들(304)은 정적인 객체들(302)과 함께 렌더링 될 때 정적인 객체들(302)과 연관된 오디오를 강화하는 강화들을 포함할 수 있다. 예를 들어, 동적인 객체들(304)은 렌더기가 정적인 객체들(302)을 강화하는데 이용할 수 있는 음향 심리학적 정보를 포함할 수 있다. 동적인 객체들(304)은 또한 렌더기가 오디오 장면을 강화하는데 이용할 수 있는 배경 객체들(예를 들어, 지나가는 비행기)을 포함할 수 있다. 그러나 동적인 객체들(304)은 배경 객체들을 요구하지 않는다. 동적인 객체들(304)은 다이얼로그 또는 임의의 다른 오디오 객체를 포함할 수 있다.

- [0050] 정적인 객체들(302)과 연관된 메타데이터는 거의 없거나 존재하지 않을 수 있다. 일 실시예에서, 이 메타데이터는 단순히 어느 채널이 정적인 객체들(302)에 대응하는지를 표시하는 "채널"의 객체 속성을 포함한다. 이 메타데이터는 몇몇 실시예들에서 변경되지 않기 때문에, 이에 따라 정적인 객체들(302)은 그들의 객체 속성들 면에서 정적이다. 대조적으로, 동적인 객체들(304)은 변하는 위치, 속도 등과 같이 변하는 객체 속성들을 포함할 수 있다. 따라서, 이 객체들(304)과 연관된 메타데이터는 동적일 수 있다. 그러나 몇몇 상황들에서, 정적인 객체들(302)과 연관된 메타데이터는 시간에 경과함에 따라 변할 수 있는 반면에, 동적인 객체들(304)과 연관된 메타데이터는 동일함을 유지할 수 있다.
- [0051] 또한, 위에서 언급한 바와 같이, 몇몇의 동적인 객체들(304)은 오디오 페이로드를 거의 또는 전혀 포함하지 않을 수 있다. 예를 들어, 환경 객체들(304)은 장면이 발생하는 음향 환경의 원하는 특성들을 특정할 수 있다. 이 동적인 객체들(304)은 방, 사무실, 대성당, 스타디움 등과 같이 오디오 장면이 발생하는 빌딩 또는 옥외 영역의 타입에 관한 정보를 포함할 수 있다. 렌더기는 예를 들어, 표시되는 환경에 대응하는 적절한 양의 반향(reverberation) 또는 지연을 적용함으로써 정적인 객체들(302)의 오디오의 재생을 조정하는데 이 정보를 이용할 수 있다. 환경적인 동적인 객체들(304)은 또한 몇몇 구현들에서 오디오 페이로드를 포함할 수 있다. 환경 객체들의 몇몇 예들은 도 4에 대해서 후술된다.
- [0052] 메타데이터를 포함할 수 있지만 페이로드를 전혀 또는 거의 포함하고 있지 않은 다른 타입의 객체는 오디오 한정 객체(audio definition object)이다. 일 실시예에서, 사용자 시스템은 오디오 한정 객체들의 수신 시에 렌더기에 의해 렌더링될 수 있는 오디오 클립들 또는 사운드들의 라이브러리(library)를 포함할 수 있다. 오디오 한정 객체는 클립을 얼마나 오래 재생할지, 클립을 되풀이해서 재생할지(loop) 여부 등에 대한 명령들과 함께 사용자 시스템상에 저장된 오디오 클립 또는 사운드에 대한 참조(reference)를 포함할 수 있다. 오디오 스트림은 오디오 한정 객체들로부터 부분적으로 또는 심지어 단독으로 구성될 수 있으며, 실제 오디오 데이터의 일부 또는 모두는 사용자 시스템 상에 저장된다(또는 다른 서버로부터 액세스 가능하게 됨). 다른 실시예에서, 스트리밍 모듈(322)은 사용자 시스템에 복수의 오디오 한정 객체들을 송신하고, 후속하여 복수의 오디오 페이로드 객체들을 송신하여 메타데이터와 실제 오디오를 분리시킨다. 다수의 다른 구성들이 가능하다.
- [0053] 콘텐츠 생성기들은 서술적인 컴퓨터 언어(예를 들어, 오디오 객체 생성 시스템(110)을 이용하여)를 이용하여 정적인 객체들(302) 또는 동적인 객체들(304)을 선언(declare)한다. 추후에 스트리밍될 오디오 콘텐츠를 생성할 때, 콘텐츠 생성기는 원하는 수의 정적인 객체들(302)을 선언할 수 있다. 예를 들어, 콘텐츠 생성기는 다이얼로그 정적인 객체(302)(예를 들어, 중앙 채널에 대응함) 또는 임의의 다른 수의 정적인 객체들(302)이 항상 온(on)이 되도록 요청할 수 있다. 이러한 "항상 온(always on)" 특성은 또한 정적인 객체들(302)을 정적이 되게 할 수 있다. 대조적으로, 동적인 객체들(304)은 왕래(come and go)할 수 있으며, 오디오 스트림에 항상 존재하는 것은 아닐 수 있다. 물론, 이 특징들은 반전될 수 있다. 예를 들어, 정적인 객체들(302)을 게이팅(gate) 또는 그렇지 않고 토글링(toggle)하는 것이 바람직할 수 있다. 예를 들어, 다이얼로그가 주어진 정적인 객체(302)에 존재하지 않을 때, 오디오 스트림에 그 정적인 객체(302)를 포함시키지 않는 것은 컴퓨팅 및 네트워크 자원들을 절약할 수 있다.
- [0054] 도 4는 객체-지향 오디오 스트리밍 포맷(400)의 실시예를 예시한다. 오디오 스트리밍 포맷은 상술된 비트 스트림들 중 임의의 비트 스트림에 대응할 수 있는 비트 스트림(414)을 포함한다. 비트 스트림(414)의 포맷(400)은 연속적으로 더욱 상세한 형상들(420, 430)로 분해된다. 도시된 비트 스트림 포맷(400)은 단순히 예시적인 실시예이며, 구현에 의존하여 달라질 수 있다.
- [0055] 도시된 실시예에서, 비트 스트림(414)은 스트림 헤더(412) 및 매크로 프레임들(420)을 포함한다. 스트림 헤더(412)는 비트 스트림(414)의 선두 또는 말단에 발생할 수 있다. 스트림 헤더(412)에 포함될 수 있는 정보의 몇몇 예들은 스트림의 창시자, 스트림의 기원, 저작권 정보, 스트림의 생성 및/또는 전달에 관련된 타임스탬프, 스트림의 길이, 스트림을 인코딩하는데 어느 코덱이 사용되었는지에 관한 정보 등을 포함한다. 스트림 헤더(412)는 스트림(414)을 적절히 디코딩하기 위해 디코더 및/또는 렌더기에 의해 이용될 수 있다.
- [0056] 매크로 프레임들(420)은 비트 스트림(414)을 데이터의 섹션들로 분할한다. 각 매크로 프레임(420)은 오디오의 시간 슬라이스(time slice) 또는 오디오 장면에 대응할 수 있다. 각 매크로 프레임(420)은 매크로 프레임 헤더(422) 및 개별적인 프레임들(430)을 추가로 포함한다. 매크로 프레임 헤더(422)는 매크로 프레임에 포함된 오디오 객체들의 수, 매크로 프레임(420)에 대응하는 타임스탬프 등을 정의할 수 있다. 몇몇 구현들에서, 매크로 프레임 헤더(422)는 매크로 프레임(420)의 프레임들(430) 이후에 위치될 수 있다. 개별적인 프레임들(430)은 단일의 오디오 객체를 각각 나타낸다. 그러나 프레임들(430)은 또한 몇몇 구현들에서 다수의 오디오 객체들을 나타낼 수 있다. 일 실시예에서, 렌더기는 매크로 프레임(420)과 연관된 오디오 객체들을 렌더링하

기 이전에 전체 매크로 프레임(420)을 수신한다.

[0057] 각 프레임(430)은 객체 메타데이터를 포함하는 프레임 헤더(432) 및 오디오 페이로드(434)를 포함한다. 몇몇 구현들에서, 프레임 헤더(432)는 오디오 페이로드(434) 이후에 위치될 수 있다. 그러나 상술한 바와 같이, 몇몇의 오디오 객체들은 단지 메타데이터(432) 또는 단지 오디오 페이로드(434) 중 어느 하나만을 가질 수 있다. 따라서, 몇몇 프레임들(432)은 객체 메타데이터가 거의 없거나 전혀 없는(또는 헤더가 전혀 없음) 프레임 헤더(432)를 포함할 수 있고, 몇몇 프레임들(432)은 오디오 페이로드(434)를 거의 또는 전혀 포함하지 않을 수 있다.

[0058] 프레임 헤더(432)의 객체 메타데이터는 객체 속성들에 관한 정보를 포함할 수 있다. 다음의 표들은 객체 속성들을 정의하는데 이용될 수 있는 메타데이터의 예들을 예시한다. 특히, 표 1은 속성 명칭 및 속성 설명에 의해 구성되는 다양한 객체 속성들을 예시한다. 도시된 속성들보다 더 적은 또는 더 많은 속성들이 몇몇 설계들에서 구현될 수 있다.

표 1

| 속성 명칭              | 속성 설명  |
|--------------------|--|
| ENABLE_PROCESS     | 모든 프로세스들을 인에이블/디스에이블, 모든 소스들에 적용             |
| ENABLE_3D_POSITION | 3D 위치 프로세스를 인에이블/디스에이블                       |
| SRC_X              | 사운드 소스의 X축 위치를 수정, 이는 청취자 및/또는 카메라에 관련됨      |
| SRC_Y              | 사운드 소스의 Y축 위치를 수정, 이는 청취자 및/또는 카메라에 관련됨      |
| SRC_Z              | 사운드 소스의 Z축 위치를 수정, 이는 청취자 및/또는 카메라에 관련됨      |
| ENABLE_DOPPLER     | 도플러(Doppler) 프로세스를 인에이블/디스에이블                |
| DOPPLER_FACT       | 도플러 피치 효과(Doppler pitch effect)의 스케일링/과장을 허용 |
| SRC_VEL_X          | X축 방향의 사운드 소스의 속도를 수정                        |
| SRC_VEL_Y          | Y축 방향의 사운드 소스의 속도를 수정                        |
| SRC_VEL_Z          | Z축 방향의 사운드 소스의 속도를 수정                        |
| ENABLE_DISTANCE    | 거리 감쇠 프로세스를 인에이블/디스에이블                       |
| MINIMUM_DIST       | 거리 감쇠가 신호를 감쇠하기 시작하는 청취자로부터의 거리              |
| MAXIMUM_DIST       | 거리 감쇠가 더이상 신호를 감쇠하지 않는 청취자로부터의 거리            |
| SILENCE_AFT_MAX    | 최대 거리에 도달한 이후 신호를 침묵시킴                       |
| ROLLOFF_FACT       | 사운드 신호 레벨이 청취자로부터의 거리의 함수로서 쇠퇴하는 레이트         |
| LISTENER_RELATIVE  | 소스 위치가 고정되거나 또는 카메라가 아닌 청취자에 관련되는지 여부를 설정    |
| LISTENER_X         | X-축에 따른 청취자의 위치                              |
| LISTENER_Y         | Y-축에 따른 청취자의 위치                              |
| LISTENER_Z         | Z-축에 따른 청취자의 위치                              |
| LISTENER_VEL_X     | X-축에 따른 청취자의 속도                              |
| LISTENER_VEL_Y     | Y-축에 따른 청취자의 속도                              |
| LISTENER_VEL_Z     | Z-축에 따른 청취자의 속도                              |
| ENABLE_ORIENTATION | 청취자 지향성 관리자를 인에이블/디스에이블(이는 모든 소스들에 적용함)      |
| LISTENER_ABOVE_X   | 청취자 위의 X-축 지향성 벡터                            |
| LISTENER_ABOVE_Y   | 청취자 위의 Y-축 지향성 벡터                            |

[0059]

|                        |   |
|------------------------|---|
| LISTENER_ABOVE_Z       | 청취자 위의 Z-축 지향성 벡터   |
| LISTENER_FRONT_X       | 청취자 앞의 X-축 지향성 벡터   |
| LISTENER_FRONT_Y       | 청취자 앞의 Y-축 지향성 벡터   |
| LISTENER_FRONT_Z       | 청취자 앞의 Z-축 지향성 벡터   |
| ENABLE_MACROSCOPIC     | 객체의 거시적인 규격(macroscopic specification)의 이용을 인에이블 또는 디스에이블                                   |
| MACROSCOPIC_X          | 사운드 방출의 x 치수 크기를 규정   |
| MACROSCOPIC_Y          | 사운드 방출의 y 치수 크기를 규정   |
| MACROSCOPIC_Z          | 사운드 방출의 z 치수 크기를 규정   |
| ENABLE_SRC_ORIENT      | 소스 상에서 지향성의 이용을 인에이블 또는 디스에이블   |
| SRC_FRONT_X            | 사운드 객체 앞의 X-축 지향성 벡터  |
| SRC_FRONT_Y            | 사운드 객체 앞의 Y-축 지향성 벡터  |
| SRC_FRONT_Z            | 사운드 객체 앞의 Z-축 지향성 벡터  |
| SRC_ABOVE_X            | 사운드 객체 위의 X-축 지향성 벡터  |
| SRC_ABOVE_Y            | 사운드 객체 위의 Y-축 지향성 벡터  |
| SRC_ABOVE_Z            | 사운드 객체 위의 Z-축 지향성 벡터  |
| ENABLE_DIRECTIVITY     | 배향성 프로세스를 인에이블 또는 디스에이블   |
| DIRECTIVITY_MIN_ANGLE  | 배향성 감쇠를 위해 360도로 정규화된 최소각을 설정<br>이 각은 원뿔(cone)을 생성하는 소스의 전면 지향성을 중심으로 집중됨                   |
| DIRECTIVITY_MAX_ANGLE  | 지향성 감쇠를 위해 360도로 정규화된 최대각을 설정   |
| DIRECTIVITY_REAR_LEVEL | 전체-스케일 중 규정된 소량만큼 신호를 감쇠  |
| ENABLE_OBSTRUCTION     | 폐색 프로세스를 인에이블 또는 디스에이블  |
| OBSTRUCT_PRESET        | 프리셋 HF 레벨/ 레벨 설정(아래의 표 2 참조)  |
| REVERB_ENABLE_PROCSS   | 반향 프로세스를 인에이블/디스에이블(모든 소스들에 영향을 미침)   |
| REVERB_DECAY           | 60dB의 쇠퇴를 위해 반향 신호에 대한 시간을 선택(전체적인 프로세스)  |
| REVERB_MIX             | 이용할 프로세싱된 신호로 원래의 신호의 양을 규정   |
| REVERB_PRESET          | 환경에 기초하여 미리 규정된 반향 구성을 선택<br>이는 변경시에 쇠퇴 시간을 수정할 수 있음<br>몇개의 미리 규정된 프리셋들이 이용 가능함(아래의 표 3 참조) |

[0060]

[0061]

[0062]

표 1- 예시적인 객체 속성들

표 1에서 나열된 OBSTRUCT\_PRESET(폐색 프리셋)에 대한 예시적인 값들은 아래의 표 2에서 도시된다. 폐색 프리셋 값은 사운드 소스가 카메라 또는 청취자의 시점으로부터 차폐되거나 차단되는 정도에 영향을 미칠 수 있다. 따라서, 예를 들어, 두꺼운 문 뒤에서 발산하는 사운드 소스는 커튼 뒤에서 발산하는 사운드 소스와 상이하게 랜더링될 수 있다. 상술한 바와 같이, 랜더기는 이들 및 다른 객체 속성들의 값들에 기초하여 임의의 원하는 랜더링 기법(또는 전혀 없음)을 수행할 수 있다.

표 2

| 폐색 프리셋 | 타입    |
|--------|-------|
| 1      | 단일문   |
| 2      | 이중문   |
| 3      | 얇은문   |
| 4      | 두꺼운 문 |
| 5      | 나무벽   |
| 6      | 벽돌 벽  |
| 7      | 석재 벽  |
| 8      | 커튼    |

[0063]

[0064] 표 2- 예시적인 폐색 프리셋들

[0065] 폐색(때때로 차폐로서 지칭됨) 프리셋과 유사하게, REVERB\_PRESET(반향 프리셋)이 표 3에서 도시된 바와 같은 예시적인 값들을 포함할 수 있다. 이 반향 값들은 사운드 소스가 위치될 수 있는 환경들의 타입들에 대응한다. 따라서 강당에서 발산하는 사운드 소스는 거실에서 발산하는 사운드 소스와 상이하게 렌더링될 수 있다. 일 실시예에서, 환경 객체는 후술되는 것과 같은 프리셋 값들을 포함하는 반향 속성을 포함한다.

표 3

| 반향 프리셋 | 타입          |
|--------|-------------|
| 1      | 좁은길         |
| 2      | 원형경기장       |
| 3      | 강당          |
| 4      | 침실          |
| 5      | 동굴          |
| 6      | 응접실         |
| 7      | 도시          |
| 8      | 콘서트 홀       |
| 9      | 숲           |
| 10     | 복도          |
| 11     | 차고          |
| 12     | 큰 방         |
| 13     | 거실          |
| 14     | 중간 크기의 방    |
| 15     | 산           |
| 16     | 주차장         |
| 17     | 플레이트(plate) |
| 18     | 방           |
| 19     | 수중          |

[0066]

- [0067] 표 3 - 예시적인 반향 프리셋들
- [0068] 몇몇 실시예들에서, 환경 객체들은 단순히 상술한 반향 프리셋들을 이용하여 기술되지 않는다. 대신, 환경 객체들은 반향의 양(프리셋될 필요가 없음), 에코(echo)의 양, 배경 잡음의 정도 등과 같은 하나 이상의 속성들로 기술될 수 있다. 다수의 다른 구성들이 가능하다. 유사하게, 오디오 객체들의 속성들은 일반적으로 값들이 아닌 형태들을 가질 수 있다. 예를 들어, 속성은 사운드 소스의 작용 또는 특성을 정의하는 코드 또는 명령들의 단편(snippet)을 포함할 수 있다.
- [0069] 도 5a는 오디오 스트림 어셈블리 프로세스(500A)의 실시예를 예시한다. 오디오 스트림 어셈블리 프로세스(500A)는 여기서 기술된 시스템들 중 임의의 시스템에 의해 구현될 수 있다. 예를 들어, 오디오 스트림 어셈블리 프로세스(500A)는 상술한 객체-지향 인코더들 또는 스트리밍 모듈들 중 임의의 것에 의해 구현될 수 있다. 스트리밍 어셈블리 프로세스(500A)는 적어도 하나의 오디오 객체로부터의 오디오 스트림을 어셈블리한다.
- [0070] 블록(502)에서, 오디오 객체가 스트림에 대해 선택된다. 오디오 객체는 상술한 오디오 객체 생성 모듈(110)에 의해 생성될 수 있다. 그럼으로써, 오디오 객체를 선택하는 것은 객체 데이터 저장소(116)의 오디오 객체에 액세스하는 것을 포함할 수 있다. 대안적으로, 스트리밍 모듈(122)은 컴퓨터 스토리지로부터 오디오 객체에 액세스할 수 있다. 예시의 편의를 위해, 이 예시적인 도면은 단일의 객체를 스트리밍하는 것을 기술하지만, 다수의 객체들이 오디오 스트림으로 스트리밍될 수 있다는 것이 이해되어야 한다. 선택된 객체는 정적인 객체 또는 동적인 객체일 수 있다. 특정한 예에서, 선택된 객체는 메타데이터 및 오디오 페이로드를 갖는다.
- [0071] 객체의 메타데이터를 갖는 객체 헤더가 블록(504)에서 어셈블리된다. 이 메타데이터는 객체 속성들의 임의의 설명을 포함할 수 있으며, 이들의 몇몇 예들이 상술되었다. 블록(506)에서, 객체의 오디오 신호 데이터를 갖는 오디오 페이로드가 제공된다.
- [0072] 객체 헤더 및 오디오 페이로드는 블록(508)에서 오디오 스트림을 형성하도록 조합된다. 오디오 스트림을 형성하는 것은 오디오 스트림을 인코딩하는 것, 오디오 스트림을 압축하는 것 등을 포함할 수 있다. 블록(510)에서, 오디오 스트림은 네트워크를 통해 전송된다. 오디오 스트림이 임의의 스트리밍 기법을 이용하여 스트리밍될 수 있지만, 오디오 스트림은 또한 사용자 시스템에 업로딩(또는, 역으로 사용자 시스템에 의해 다운로드)될 수 있다. 그 후, 오디오 시스템은 도 5b에 관하여 후술되는 바와 같이 사용자 시스템에 의해 렌더링될 수 있다.
- [0073] 도 5b는 오디오 스트림 렌더링 프로세스(500B)의 실시예를 예시한다. 오디오 스트림 렌더링 프로세스(500B)는 여기서 기술된 시스템들 중 임의의 시스템에 의해 구현될 수 있다. 예를 들어, 오디오 스트림 렌더링 프로세스(500B)는 여기서 기술된 렌더기들 중 임의의 렌더기에 의해 구현될 수 있다.
- [0074] 블록(522)에서, 객체-지향 오디오 스트림이 수신된다. 이 오디오 스트림은 프로세스(500A)의 기법들을 이용하여 또는 상술된 다른 기법들을 통해 생성될 수 있다. 오디오 스트림의 객체 메타데이터는 블록(524)에서 액세스된다. 이 메타데이터는 예를 들어, 스트림을 인코딩하는데 이용된 동일한 코덱을 이용하여 스트림을 디코딩함으로써 획득될 수 있다.
- [0075] 메타데이터의 하나 이상의 객체 속성들은 블록(526)에서 식별된다. 이 객체 속성들의 값들은 스트림의 오디오 객체들을 렌더링하기 위한 단서들로서 렌더기에 의해 식별될 수 있다.
- [0076] 오디오 스트림의 오디오 신호는 블록(528)에서 렌더링된다. 도시된 실시예에서, 오디오 스트림은 출력 오디오를 생성하기 위해 하나 이상의 객체 속성들에 따라 렌더링된다. 출력 오디오는 블록(530)에서 하나 이상의 로드스피커들에 공급된다.
- [0077] IV. 적응형 스트리밍 및 렌더링 실시예들
- [0078] 적응형 스트리밍 모듈(122B) 및 적응형 렌더기(142B)는 도 1b에 관하여 상술되었다. 적응형 스트리밍 모듈(622) 및 적응형 렌더기(642)의 보다 상세한 실시예들이 도 6의 시스템(600)에서 도시된다.
- [0079] 도 6에서, 적응형 스트리밍 모듈(622)은 우선순위 모듈(625), 네트워크 자원 모니터(626), 객체-지향 인코더(612), 오디오 통신 모듈(628)을 포함하는 몇 개의 컴포넌트들을 갖는다. 적응형 렌더기(642)는 컴퓨팅 자원

모니터(644) 및 렌더링 모듈(646)을 포함한다. 도시된 컴포넌트들 중 몇 개의 컴포넌트들은 상이한 구현들에서 생략될 수 있다. 객체-지향 인코더(612)는 상술한 인코딩 특징들 중 임의의 특징들을 포함할 수 있다. 오디오 통신 모듈(628)은 네트워크(도시되지 않음)를 통해 비트 스트림(614)을 적응형 렌더기(642)에 전송할 수 있다.

[0080] 우선순위 모듈(624)은 우선순위 값들 또는 다른 우선순위 정보를 오디오 객체들에 적용할 수 있다. 일 실시예에서, 각각의 객체는 숫자 값 동일 수 있는 우선순위 값을 가질 수 있다. 우선순위 값들은 렌더링 견지(rendering standpoint)로부터 객체들의 상대적 중요성을 표시할 수 있다. 더 높은 우선순위를 갖는 객체들은 더 낮은 우선순위의 객체들보다 렌더링하는데 더 중요하게 될 수 있다. 따라서, 자원들이 제한되는 경우, 상대적으로 더 낮은 우선순위를 갖는 객체들은 무시될 수 있다. 우선순위는 상술한 오디오 객체 생성 시스템들(110)을 이용하여 콘텐츠 생성기에 의해 초기에 설정될 수 있다.

[0081] 예로서, 비디오에 대한 다이얼로그를 포함하는 다이얼로그 객체는 배경 사운드 객체보다 상대적으로 더 높은 우선순위를 가질 수 있다. 예를 들어, 우선순위 값들이 1 내지 5의 등급 상에 있는 경우, 다이얼로그 객체는 1의 우선순위 값(최고 우선순위를 의미함)을 갖는 반면에, 배경 사운드 객체는 더 낮은 우선순위(예를 들어, 2 내지 5 중 임의의 등급)를 가질 수 있다. 우선순위 모듈(624)은 특정한 우선순위 레벨들을 만족하는 객체들을 전송하기 위한 임계치들을 설정할 수 있다. 예를 들어, 우선순위 모듈(624)은 3의 임계치를 설정하여, 1, 2, 및 3의 우선순위를 갖는 객체들이 사용자 시스템에 전송되게 하는 반면에, 4 또는 5의 우선순위를 갖는 객체는 전송되지 않게 할 수 있다.

[0082] 우선순위 모듈(624)은 네트워크 자원 모니터(626)에 의해 결정된 바와 같이, 변하는 네트워크 조건들에 기초하여, 이 임계치를 동적으로 설정할 수 있다. 네트워크 자원 모니터(626)는 대역폭, 레이턴시 등과 같은 이용 가능한 네트워크 자원들 또는 다른 서비스 품질 측정들을 모니터링할 수 있다. 네트워크 자원 모니터(626)는 이 정보를 우선순위 모듈(624)에 제공할 수 있다. 이 정보를 이용하여, 우선순위 모듈(624)은 네트워크 자원들이 높은 경우 더 낮은 우선순위 객체들이 사용자 시스템에 전송되는 것을 허용하도록 임계치를 조정할 수 있다. 유사하게, 우선순위 모듈(624)은 네트워크 자원이 낮을 때 더 낮은 우선순위 객체들이 전송되는 것을 방지하도록 임계치를 조정할 수 있다.

[0083] 우선순위 모듈(624)은 또한 적응형 렌더기(642)로부터 수신된 정보에 기초하여 우선순위 임계치를 조정할 수 있다. 적응형 렌더기(642)의 컴퓨팅 자원 모듈(644)은 사용자 시스템에 연결된 스피커들의 수, 사용자 시스템의 처리 성능 등과 같은 사용자 시스템의 재생 환경의 특성들을 식별할 수 있다. 컴퓨팅 자원 모듈(644)은 제어 채널(650)을 통해 컴퓨팅 자원 정보를 우선순위 모듈(624)에 통신할 수 있다. 이 정보에 기초하여, 우선순위 모듈(624)은 컴퓨팅 자원이 높은 경우 더 높은 우선순위 객체 및 더 낮은 우선순위 객체 둘 다를, 및 컴퓨팅 자원들이 낮은 경우 더 높은 우선순위 객체들만을 송신하도록 임계치를 조정할 수 있다. 그러므로 적응형 렌더기(642)의 컴퓨팅 자원 모니터(644)는 사용자 시스템에 스트리밍되는 오디오 객체들의 양 및/또는 타입을 제어할 수 있다.

[0084] 적응형 렌더기(642)는 또한 재생 환경에 기초하여 오디오 스트림들이 렌더링되는 방법을 조정할 수 있다. 예를 들어, 사용자 시스템이 2개의 스피커들에 연결된 경우, 적응형 렌더기(642)는 2개의 스피커들 상에서 오디오 객체들을 렌더링할 수 있다. 부가적인 스피커들이 사용자 시스템에 연결되는 경우, 적응형 렌더기(642)는 부가적인 채널 상에서 오디오 객체들을 또한 렌더링할 수 있다. 적응형 렌더기(642)는 또한 하나 또는 두 개의(또는 때때로 그 이상) 스피커들 상에서 오디오 객체들을 렌더링할 때 음향 심리학적 기법들을 적용할 수 있다.

[0085] 우선순위 모듈(624)은 오디오 객체들의 우선순위를 동적으로 변경할 수 있다. 예를 들어, 우선순위 모듈(624)은 서로에 대해 상대적인 우선순위를 갖도록 객체들을 설정할 수 있다. 예를 들어, 다이얼로그 객체는 우선순위 모듈(624)에 의해 최고 우선순위 값이 지정될 수 있다. 다른 객체들의 우선순위 값들은 다이얼로그 객체의 우선순위에 상대적일 수 있다. 따라서 다이얼로그 객체가 시간 기간 동안 오디오 스트림에 존재하지 않는 경우, 다른 객체들이 상대적으로 더 높은 우선순위를 가질 수 있다.

[0086] 도 7은 적응형 스트리밍 프로세스(700)의 실시예를 예시한다. 적응형 스트리밍 프로세스(700)는 시스템(600)과 같이 상술된 시스템들 중 임의의 시스템에 의해 구현될 수 있다. 적응형 스트리밍 프로세스(700)는 스트리밍 자원들의 효율적인 이용을 용이하게 한다.

[0087] 블록들(702 내지 708)은 상술한 우선순위 모듈(624)에 의해 수행될 수 있다. 블록(702)에서, 오디오 콘텐츠에 대한 요청이 원격 컴퓨터로부터 수신된다. 사용자 시스템은 예를 들어, 이 요청을 콘텐츠 서버에 송신할

수 있다. 블록(704)에서, 원격 컴퓨터 시스템의 자원들에 관한 컴퓨팅 자원 정보가 수신된다. 이 컴퓨팅 자원 정보는 사용자 시스템의 다양한 이용 가능한 자원들을 기술할 수 있고, 오디오 콘텐츠 요청과 함께 제공될 수 있다. 이용 가능한 네트워크 자원들에 관한 네트워크 자원 정보는 또한 블록(726)에서 수신된다. 이 네트워크 자원 정보는 네트워크 자원 모니터(626)에 의해 획득될 수 있다.

[0088] 우선순위 임계치는 컴퓨터 및/또는 네트워크 자원 정보에 적어도 부분적으로 기초하여 블록(708)에서 설정된다. 일 실시예에서, 우선순위 모듈(624)은 컴퓨팅 및 네트워크 자원들 둘 다가 상대적으로 높을 때 더 낮은 임계치(예를 들어, 스트림의 더 낮은 우선순위 객체들을 허용하기 위해)를 설정한다. 우선순위 모듈(624)은 컴퓨팅 또는 네트워크 자원들 중 어느 하나가 상대적으로 낮을 때 더 높은 임계치(예를 들어, 스트림의 더 높은 우선순위 객체들을 허용하기 위해)를 설정할 수 있다.

[0089] 블록들(710 내지 714)은 객체-지향 인코더(612)에 의해 수행될 수 있다. 요청된 오디오 콘텐츠의 주어진 객체에 대해서, 판단 블록(710)에서, 그 객체에 대한 우선순위 값이 이전에 설정된 임계치를 만족하는지를 결정한다. 만약 만족한다면, 객체는 오디오 스트림에 추가된다. 그렇지 않으면, 객체는 오디오 스트림에 추가되지 않고, 그에 의해 특정한 실시예들에서 네트워크 및/또는 컴퓨팅 자원들을 유리하게 절감한다.

[0090] 블록(714)에서 스트림에 부가를 위해 고려될 추가적인 객체들이 남아있는지를 추가로 결정한다. 만약 남아있다면, 프로세스(700)는 블록(710)으로 되돌아가 반복(loop)한다. 그렇지 않다면, 오디오 스트림은 예를 들어, 오디오 통신 모듈(628)에 의해 블록(716)에서 원격 컴퓨터 시스템에 전송된다.

[0091] 프로세스(700)는 공중에서 오디오 스트림을 어셈블링하는 대신에 사전-인코딩된 오디오 스트림으로부터 객체들을 제거하도록 몇몇 구현들에서 수정될 수 있다. 예를 들어, 블록(710)에서, 주어진 객체가 임계치를 만족하지 않는 우선순위를 갖는 경우, 블록(712)에서 객체는 오디오 스트림으로부터 제거될 수 있다. 따라서 콘텐츠 생성기들은 다양한 객체들을 갖는 오디오 스트림을 콘텐츠 서버에 제공할 수 있고, 콘텐츠 서버의 적응형 스트리밍 모듈은 객체들의 우선순위에 기초하여 객체들 중 몇몇의 객체들을 동적으로 제거할 수 있다. 그러므로 스트리밍을 위한 오디오 객체들을 선택하는 것은 스트림에 객체를 추가하는 것 또는 스트림으로부터 객체들을 제거하는 것, 또는 둘 다를 포함한다.

[0092] 도 8은 적응형 렌더링 프로세스(800)의 실시예를 예시한다. 적응형 렌더링 프로세스(800)는 시스템(600)과 같이 상술된 시스템들 중 임의의 시스템에 의해 구현될 수 있다. 적응형 렌더링 프로세스(800)는 또한 스트리밍 자원들의 효율적인 이용을 용이하게 한다.

[0093] 블록(802)에서, 복수의 오디오 객체들을 갖는 오디오 스트림이 사용자 시스템의 렌더기에 의해 수신된다. 예를 들어, 적응형 렌더기(642)는 오디오 객체들을 수신할 수 있다. 재생 환경 정보는 블록(804)에서 액세스된다. 재생 환경 정보는 적응형 렌더기(642)의 컴퓨팅 자원 모니터(644)에 의해 액세스될 수 있다. 이 자원 정보는 스피커 구성들, 컴퓨팅 전력 등에 관한 정보를 포함할 수 있다.

[0094] 블록들(806 내지 810)은 적응형 렌더기(642)의 렌더링 모듈(646)에 의해 구현될 수 있다. 블록(806)에서, 하나 이상의 오디오 객체들은 환경 정보에 적어도 부분적으로 기초하여 선택된다. 렌더링 모듈(646)은 렌더링할 객체들을 선택하기 위해 객체들의 우선순위 값들을 이용할 수 있다. 다른 실시예에서, 렌더링 모듈(646)은 우선순위 값에 기초하여 객체들을 선택하지 않고, 대신에, 더 적은 스피커 채널들로 객체들을 다운-믹스하거나, 그렇지 않으면 오디오를 렌더링하기 위해 더 적은 프로세싱 자원들을 이용한다. 오디오 객체들은 블록(808)에서 출력 오디오를 생성하기 위해 렌더링된다. 렌더링되는 오디오는 블록(810)에서 하나 이상의 스피커들로 출력된다.

[0095] V. 오디오 객체 생성 실시예들.

[0096] 도 9 내지 도 11은 영화들, 텔레비전, 팟캐스팅(podcast) 등과 같이 오디오-비주얼 채현들의 문맥에서 예시적인 오디오 객체 생성 기법들을 기술한다. 그러나 도 9 내지 도 11에 관하여 기술되는 특징들 중 일부 또는 모두는 또한 순(pure) 오디오 문맥(예를 들어, 수반하는 비디오 없이)에서 구현될 수 있다.

[0097] 도 9는 객체-지향 오디오 캡처를 위한 예시적인 장면(900)을 예시한다. 장면(900)은 예를 들어, 영화, 텔레비전, 또는 다른 비디오를 위해 구성될 수 있는 오디오-비디오 장면의 단순화된 뷰를 나타낸다. 장면(900)에서, 2명의 배우들(910)이 공연중이며 그들의 사운드들 및 행동들은 마이크로폰(920) 및 카메라(930)에 의해 각각 기록된다. 몇몇 경우들에서, 배우들(910)이 개별적인 마이크로폰들을 착용할 수 있지만, 단순함을 위해, 단일의 마이크로폰(920)이 예시된다. 유사하게, 개별적인 마이크로폰들이 또한 프롭들(도시되지 않



음)을 위해 공급될 수 있다.

- [0098] 제공된 장면(900)에서 사운드 소스들(예를 들어, 배우들)의 위치, 속도, 및 다른 속성들을 결정하기 위해, 위치-트래킹 디바이스들(912)가 제공된다. 이 위치-트래킹 디바이스들(912)은 GPS 디바이스들, 모션 캡처 수트들, 레이저 거리 측정기(laser range finder)들 등을 포함할 수 있다. 위치-트래킹 디바이스들(912)로부터의 데이터는 마이크로폰(920)(또는 마이크로폰들)으로부터의 데이터와 함께 오디오 객체 생성 시스템(100)에 전송될 수 있다. 위치-트래킹 디바이스들(912)로부터의 데이터에 포함된 타임스탬프들은 오디오의 각 인스턴스(instance)에 대한 위치 데이터를 제공하도록 마이크로폰(920) 및/또는 카메라(930)로부터 획득한 타임스탬프들과 상관될 수 있다. 이 위치 데이터는 위치 속성을 갖는 오디오 객체들을 생성하는데 이용될 수 있다. 유사하게, 속도 데이터는 위치-트래킹 디바이스들(912)로부터 획득될 수 있거나 또는 위치 데이터로부터 유도될 수 있다.
- [0099] 위치-트래킹 디바이스들(912)로부터의 위치 데이터(예를 들어, GPS-유도된 위도 및 경도)는 위치 데이터로서 직접 이용될 수 있거나 또는 좌표계로 번역될 수 있다. 예를 들어, 3차원들(x, y, 및 z)의 데카르트 좌표들(940)이 오디오 객체 위치를 트래킹하는데 이용될 수 있다. 구 좌표들 또는 원기둥 좌표와 같이, 데카르트 좌표들이 아닌 좌표계들이 또한 이용될 수 있다. 좌표계(940)의 기원은 일 실시예에서 카메라(930)일 수 있다. 이 어레인지먼트(arrangement)를 용이하게 하기 위해, 카메라(930)는 또한 오디오 객체들에 상대적인 자신의 위치를 결정하도록 위치-트래킹 디바이스(912)를 또한 포함할 수 있다. 따라서, 카메라(930)의 위치가 변하는 경우조차도, 장면(900)에서 오디오 객체들의 위치는 카메라(930)의 위치에 여전히 상대적으로 될 수 있다.
- [0100] 위치 데이터는 또한 오디오-비주얼 제작물의 촬영 후 편집(post-production) 동안 오디오 객체들에 적용될 수 있다. 애니메이션 제작물들의 경우, 애니메이션되는(animated) 객체들(예를 들어, 캐릭터들)의 좌표들은 콘텐츠 생성기들에게 알려질 수 있다. 이 좌표들은 오디오 객체들을 생성하기 위해 각각의 애니메이션되는 객체들에 의해 생성된 오디오와 자동으로 연관될 수 있다.
- [0101] 도 10은 도 9에 관하여 상술한 특징들을 구현할 수 있는 객체-지향 오디오 캡처를 위한 시스템(1000)을 개략적으로 예시한다. 시스템(1000)에서, 사운드 소스 위치 데이터(1002) 및 마이크로폰 데이터(1006)가 객체 생성 모듈(1014)에 제공된다. 객체 생성 모듈(1014)은 상술된 객체 생성 모듈들(114A, 114B)의 모든 특징들을 포함할 수 있다. 객체 생성 모듈(1014)은 도 9에 관하여 상술된 바와 같이, 타임스탬프들(1004, 1008)에 기초하여 마이크로폰 데이터(1006)와 주어진 사운드 소스에 대한 사운드 소스 위치 데이터(1002)를 상관시킬 수 있다.
- [0102] 부가적으로, 객체 생성 모듈(1014)은 링크이거나 또는 다른 방식으로 객체들을 서로 연관시킬 수 있는 객체 링커(1020)를 포함한다. 특정한 오디오 객체들은 서로에 대해 고유하게 관련될 수 있고, 그러므로 객체 링커(1020)에 의해 서로 자동으로 링크될 수 있다. 링크된 객체들은 후술되는 것과 같은 방식으로 함께 렌더링될 수 있다.
- [0103] 객체들이 객체의 동일한 더 높은 클래스(higher class)에 관련되기 때문에, 객체들은 서로에 대해 고유하게 관련될 수 있다. 즉, 객체 생성 모듈(1014)은 부모 객체(parent object)들 및 부모 객체들의 고유한 특성들에 관련되는 자식 객체(child object)들을 포함하는 객체들의 계층들을 형성할 수 있다. 이러한 방식으로, 오디오 객체들은 컴퓨터 프로그래밍 언어들로부터 특정한 객체-지향 원리들을 차용할 수 있다. 자식 객체들을 가질 수 있는 부모 객체들의 예는 행진 밴드(marching band)이다. 행진 밴드는 트롬본들, 플루트들, 클라리넷들 등과 같이 악기들의 상이한 그룹들에 대응하는 몇 개의 섹션들을 가질 수 있다. 객체 생성 모듈(1014)을 이용하는 콘텐츠 생성기는 밴드가 부모 객체가 되고 각 섹션이 자식 객체가 되도록 지정할 수 있다. 또한, 콘텐츠 생성기는 또한 개별적인 밴드 일원들이 섹션 객체들의 자식 객체들이 되도록 지정할 수 있다. 계층 레벨들의 수를 포함하는 객체 계층의 복잡도는 콘텐츠 생성기에 의해 설정될 수 있다.
- [0104] 상술한 바와 같이, 자식 객체들은 그들의 부모 객체들의 특성들을 상속할 수 있다. 따라서, 자식 객체들은 그들의 부모 객체들의 메타데이터 중 일부 또는 모두를 상속할 수 있다. 몇몇 경우들에서, 자식 객체들은 또한 그들의 부모 객체들과 연관된 오디오 신호 데이터 중 일부 또는 모두를 상속할 수 있다. 자식 객체들은 이 메타데이터 및/또는 오디오 신호 데이터 중 일부 또는 모두를 수정할 수 있다. 예를 들어, 자식 객체는, 자식 및 부모가 위치들이 상이하지만 다른(other) 유사한 메타데이터를 갖도록 부모로부터 상속되는 위치 속성을 수정할 수 있다.
- [0105] 자식 객체들의 위치는 또한 부모 객체들의 위치로부터의 오프셋(offset)으로서 표현될 수 있거나 또는 다른

방식으로, 부모 객체의 위치로부터 유도될 수 있다. 행진 밴드의 예를 참조하면, 밴드의 섹션은 밴드의 위치로부터 오프셋되는 위치를 가질 수 있다. 밴드가 위치를 변경함에 따라, 밴드 섹션을 나타내는 자식 객체는 이 오프셋 및 부모 밴드의 위치에 기초하여 자신의 위치를 자동으로 업데이트할 수 있다. 이러한 방식으로, 상이한 위치 오프셋들을 갖는 밴드의 상이한 섹션들이 함께 이동할 수 있다.

[0106] 자식 객체와 부모 객체의 상속(inheritance)은 자식 객체와 부모 객체 사이의 공통적인 메타데이터를 초래할 수 있다. 메타데이터의 이러한 중첩은 오디오 스트림의 데이터를 최적화하거나 감소시키기 위해 상술한 객체-지향 인코더들 중 임의의 인코더에 의해 이용될 수 있다. 일 실시예에서, 부모의 메타데이터를 참조하여 중복 메타데이터를 삭제하도록 객체-지향 인코더는 자식 객체로부터 중복 메타데이터를 제거할 수 있다. 유사하게, 중복 오디오 신호 데이터가 자식 객체 및 부모 객체에 공통적인 경우, 객체-지향 인코더는 중복 오디오 신호 데이터를 감소시키거나 제거할 수 있다. 이 기법들은 객체-지향 인코더가 오디오 스트림의 중복 데이터를 감소시키거나 제거하도록 구현될 수 있는 다수의 최적화 기법들의 단순한 예들이다.

[0107] 또한, 객체 생성 모듈(1014)의 객체 링커(1020)는 자식 객체와 부모 객체를 서로 링크시킬 수 있다. 객체 링커(1020)는 두 객체들의 메타데이터에서 반영될 수 있는 두 객체들 간의 연관성을 생성함으로써 이러한 링킹(linking)을 수행할 수 있다. 객체 링커(1020)는 객체 데이터 저장소(1060)에 이러한 연관성을 저장할 수 있다. 또한, 몇몇 실시예들에서, 콘텐츠 생성기들은, 예를 들어, 객체들이 부모-자식 관계를 갖지 않는 경우조차도 서로 수동으로 객체들을 링크시킬 수 있다.

[0108] 랜더기가 2개의 링크된 객체들을 수신할 때, 랜더기는 2개의 객체들을 개별적으로 또는 함께 랜더링하도록 선택할 수 있다. 따라서, 예를 들어, 행진 밴드를 하나의 스피커를 통해 단일 지점 소스로서 랜더링하는 대신에, 랜더기는 행진 밴드를 다양한 스피커들을 통해 함께 오디오 객체들의 사운드 필드(sound field)로서 랜더링할 수 있다. 예를 들어, 밴드가 비디오에서 이동함에 따라, 랜더기는 스피커들을 통해 사운드 필드를 이동시킬 수 있다.

[0109] 보다 일반적으로, 랜더기는 다양한 방식으로 링킹 정보를 해석할 수 있다. 예를 들어, 랜더기는 서로 지연되는 상이한 시간들에 동일한 스피커, 또는 동일한 시간에 상이한 스피커들 등을 통해 링크된 객체들을 랜더링할 수 있다. 랜더기는 또한 링크된 객체들이 청취자의 머리 주위의 상이한 지점들에 있다는 느낌(impression)을 청취자에게 제공하기 위해, 음향 심리학적으로 결정된 상이한 공간적인 지점들에서 링크된 객체들을 랜더링할 수 있다. 따라서 예를 들어, 랜더기는 클라리넷 섹션이 청취자의 우측에서 행진하는 동안 트롬본 섹션이 청취자의 좌측에서 행진하는 것으로 느껴지게 할 수 있다.

[0110] 도 11은 객체-지향 오디오 캡처를 위한 프로세스(1100)의 실시예를 예시한다. 프로세스(1100)는 시스템(1000)과 같이 여기서 기술된 시스템들 중 임의의 시스템에 의해 구현될 수 있다. 예를 들어, 프로세스(1100)는 객체 생성 모듈(1014)의 객체 링커(1020)에 의해 구현될 수 있다.

[0111] 블록(1102)에서, 오디오 및 위치 데이터가 제 1 사운드 소스 및 제 2 사운드 소스에 대해 수신된다. 오디오 데이터는 마이크로폰을 이용하여 획득될 수 있는 반면에, 위치 데이터는 도 9에 관하여 상술된 기법들 중 임의의 기법을 이용하여 획득될 수 있다.

[0112] 제 1 오디오 객체는 블록(1104)에서 제 1 사운드 소스에 대해 생성된다. 유사하게, 제 2 오디오 객체는 블록(1106)에서 제 2 사운드 소스에 대해 생성된다. 블록(1108)에서 제 1 사운드 소스와 제 2 사운드 소스 간의 연관이 생성된다. 이 연관은 2개의 객체들이 객체 계층에서 관련되는지에 기초하여 객체 링커(1020)에 의해 자동으로 생성될 수 있다. 또한, 객체 링커(1020)는 임의의 2개의 유사한 속성들과 같이 객체들과 연관된 다른 메타데이터에 기초하여 자동으로 연관성을 생성할 수 있다. 이 연관은 블록(1110)에서 컴퓨터 스토리지에 저장된다.

[0113] VI. 용어

[0114] 실시예들에 의존하여, 여기서 기술된 알고리즘들 중 임의의 알고리즘의 특정한 작용들, 이벤트들, 기능들은 상이한 시퀀스로 수행될 수 있거나, 부가될 수 있거나, 병합될 수 있거나, 다함께 제외될 수 있다(예를 들어, 모든 기술된 작용들 또는 이벤트들이 알고리즘의 실행을 위해 필수적인 것은 아니다). 또한, 특정한 실시예들에서, 작용들 또는 이벤트들은 순차적인 것이 아니라, 예를 들어, 다중-스레드 처리, 인터럽트 처리, 또는 다수의 처리기를 또는 처리기들 코어들을 통해 또는 다른 병렬 아키텍처들을 통해 동시에 수행될 수 있다.

[0115] 다양한 예시적인 논리 블록들, 모듈들, 및 여기서 개시된 실시예들과 관련하여 기술되는 알고리즘 단계들은

전자 하드웨어, 컴퓨터 소프트웨어, 또는 이들 둘의 조합으로서 구현될 수 있다. 이러한 상호교환성을 명확하게 예시하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들 및 단계들은 그들의 기능성의 견지에서 일반적으로 상술되었다. 이러한 기능이 하드웨어로서 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 부과되는 설계 제약들에 의존한다. 기술된 기능은 각각의 특정한 애플리케이션에 대해서 다양한 방식으로 구현될 수 있지만, 이러한 구현 판단들은 본 개시의 범위로부터 벗어나는 것으로서 해석되어선 안 된다.

[0116] 여기서 개시된 실시예들과 관련하여 기술된 다양한 예시적인 논리 블록들 및 모듈들은 범용 처리기, 디지털 신호 처리기(DSP), 주문형 집적 회로(ASIC), 필드 프로그래밍 가능한 게이트 어레이(FPGA), 또는 다른 프로그래밍 가능한 로직 디바이스, 이산 게이트 또는 트랜지스터 로직, 이산 하드웨어 컴포넌트들, 또는 여기서 기술된 기능들을 수행하도록 설계된 이들의 임의의 조합에 의해 구현되거나 수행될 수 있다. 범용 처리기는 마이크로처리기일 수 있지만, 대안적으로 처리기는 제어기, 마이크로제어기, 또는 상태 머신, 이들의 조합들 등일 수 있다. 처리기는 또한 컴퓨팅 디바이스들의 조합, DSP 및 마이크로처리기의 조합, 복수의 마이크로처리기들, DSP와 결합된 하나 이상의 마이크로처리기들, 또는 임의의 다른 이러한 구성으로서 구현될 수 있다.

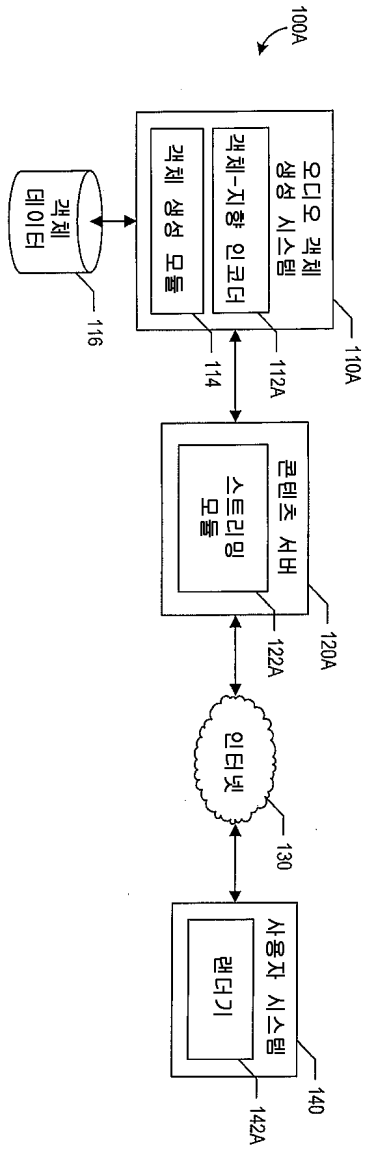
[0117] 여기서 개시된 실시예들과 관련하여 기술되는 방법, 프로세스, 또는 알고리즘의 단계들은 하드웨어로, 처리기에 의해 실행되는 소프트웨어 모듈로, 또는 이들 두 개의 조합으로 직접 구현될 수 있다. 소프트웨어 모듈은 RAM 메모리, 플래시 메모리, ROM 메모리, EPROM 메모리, EEPROM 메모리, 레지스터들, 하드 디스크들, 제거 가능한 디스크, CD-ROM, 또는 당 분야에 알려진 임의의 다른 형태의 컴퓨터-판독 가능한 저장 매체 내에 상주할 수 있다. 예시적인 저장 매체는 처리기에 결합될 수 있어서, 처리기는 저장 매체에 정보를 기록하고 이로부터 정보를 판독할 수 있게 될 수 있다. 대안적으로, 저장 매체는 처리기에 통합될 수 있다. 처리기 및 저장 매체는 ASIC 내에 상주할 수 있다. ASIC는 사용자 단말 내에 상주할 수 있다. 대안적으로, 처리기 및 저장 매체는 사용자 단말의 이산 컴포넌트들로서 상주할 수 있다.

[0118] 예를 들어, 다른 것들 중에서, "~할 수 있다("can", "might", "may"), "예를 들어," 와 같은 여기서 사용된 조건적인 언어는 일반적으로, 달리 구체적으로 언급되지 않는 한 또는 사용된 문맥 내에서 다른 방식으로 이해되지 않는 한, 특정한 실시예들이 특정한 특징들, 엘리먼트들 및/또는 상태들을 포함하는 반면에, 다른 실시예들은 이들을 포함하지 않는다는 것을 전달하도록 의도된다. 따라서 이러한 조건적인 언어는 일반적으로, 특징들, 엘리먼트들 및/또는 상태들이 포함되거나 임의의 특정한 실시예들에서 수행되든지 간에, 이 특징들, 엘리먼트들 및/또는 상태들이 어떠한 방식으로든 하나 이상의 실시예들에 대해 요구된다거나, 하나 이상의 실시예들이 창시자의 입력이나 촉진(prompting)을 통해, 또는 이들 없이, 판단을 위한 로직을 반드시 포함한다는 것을 암시하도록 의도되는 것은 아니다.

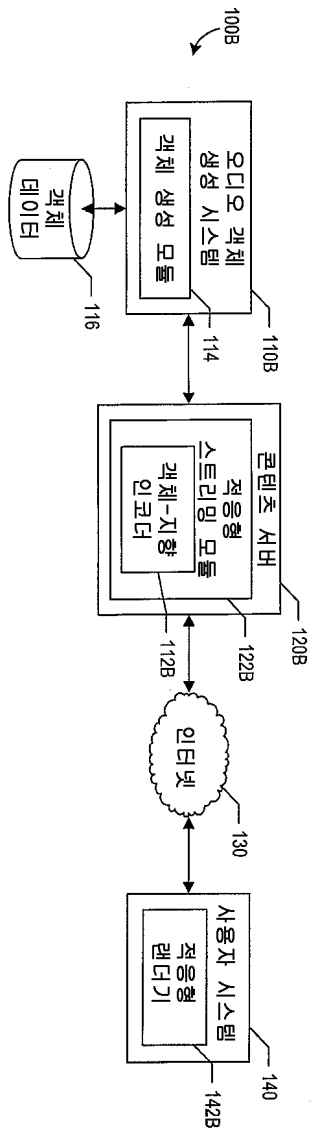
[0119] 위의 상세한 설명이 다양한 실시예들에 적용되는 신규한 특징들을 도시하고, 기술하고 지목하였지만, 예시된 알고리즘들 또는 디바이스들의 형태 및 상세들에 있어서 다양한 생략들, 대체들, 및 변경들이 본 개시의 사상으로부터 벗어남 없이 행해질 수 있다는 것이 이해될 것이다. 인식되는 바와 같이, 여기서 기술된 본 발명의 특정한 실시예들은 몇몇 특징들이 다른 것과 개별적으로 이용되거나 실시될 수 있기 때문에 여기서 기술한 특징들 및 이익들 모두를 제공하지는 않는 형태 내에서 구현될 수 있다. 여기서 개시된 특정한 발명들의 범위는 상술한 것에 의해서가 아니라 첨부된 청구항들에 의해 표시된다. 청구항들의 등가물의 취지 및 범위 내에 있게 되는 모든 변경물들은 청구범위의 범위 내로 포괄하게 된다.

도면

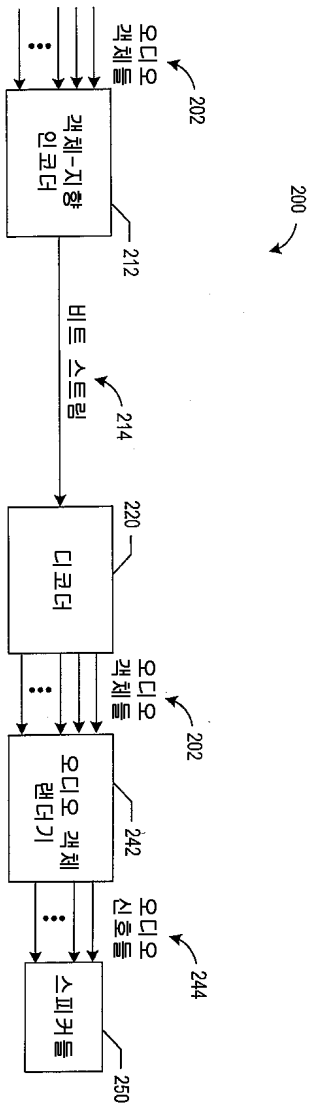
도면1a



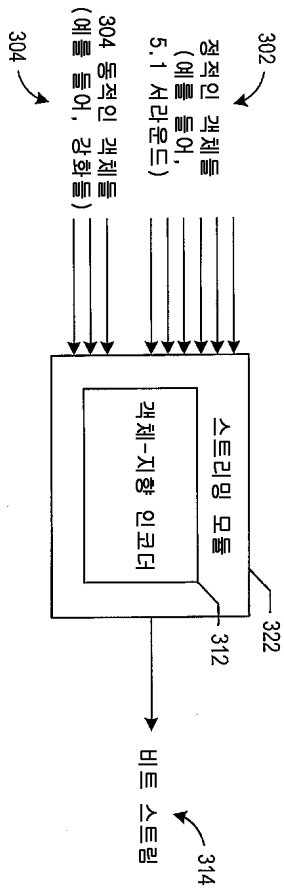
도면1b



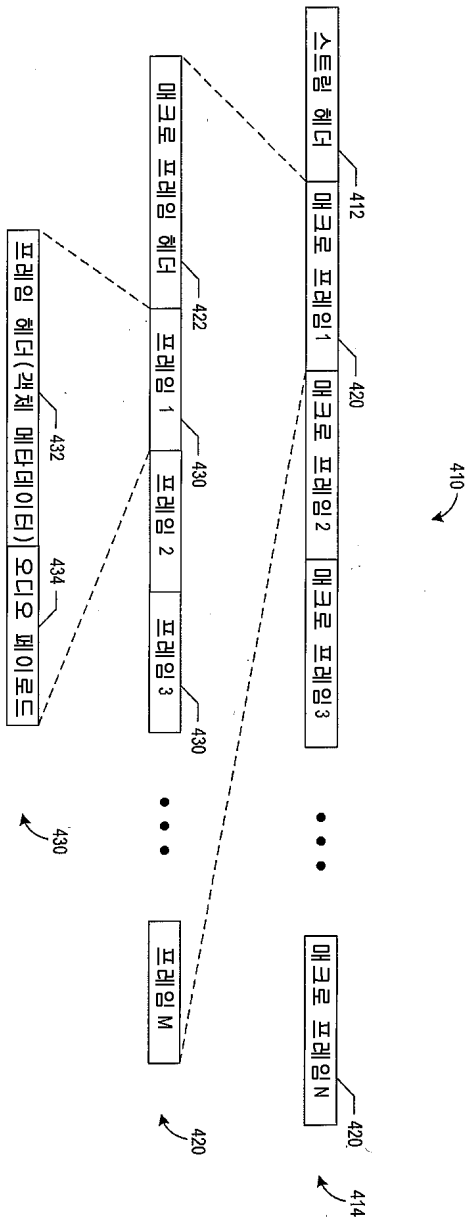
도면2



도면3

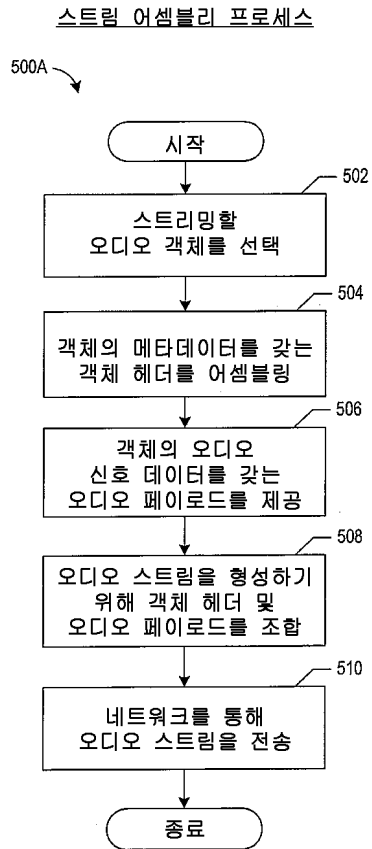


도면4

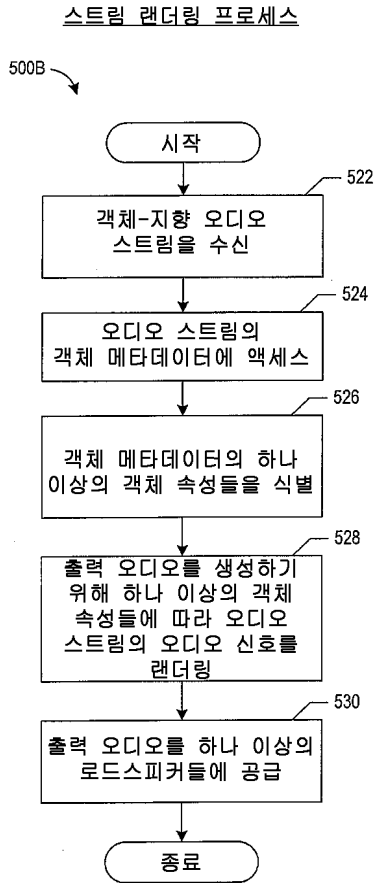




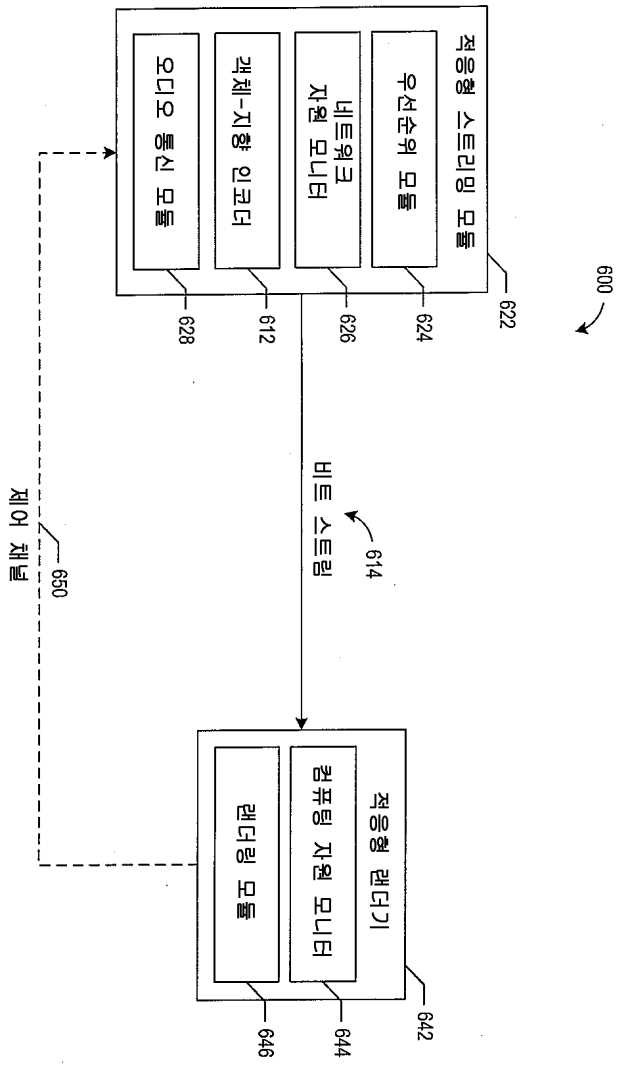
도면5a



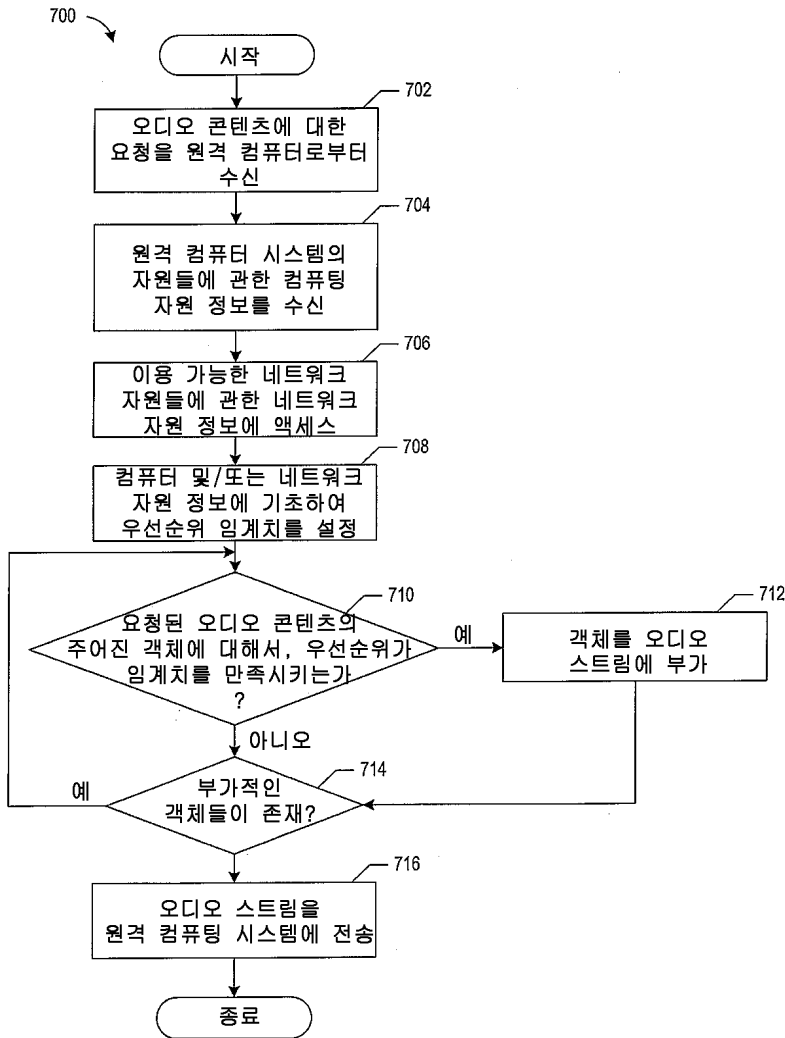
도면5b



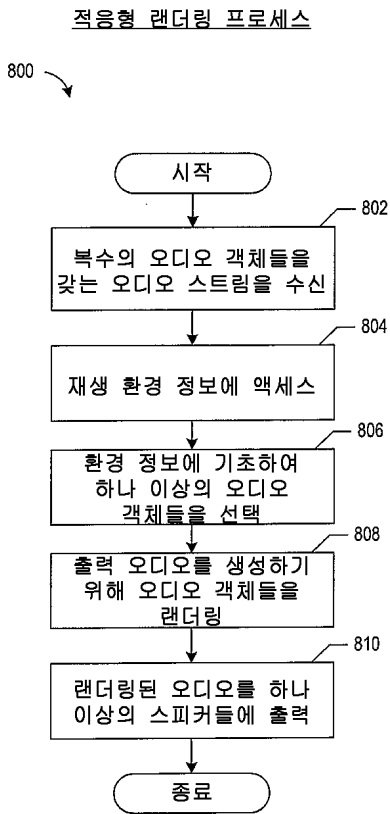
도면6



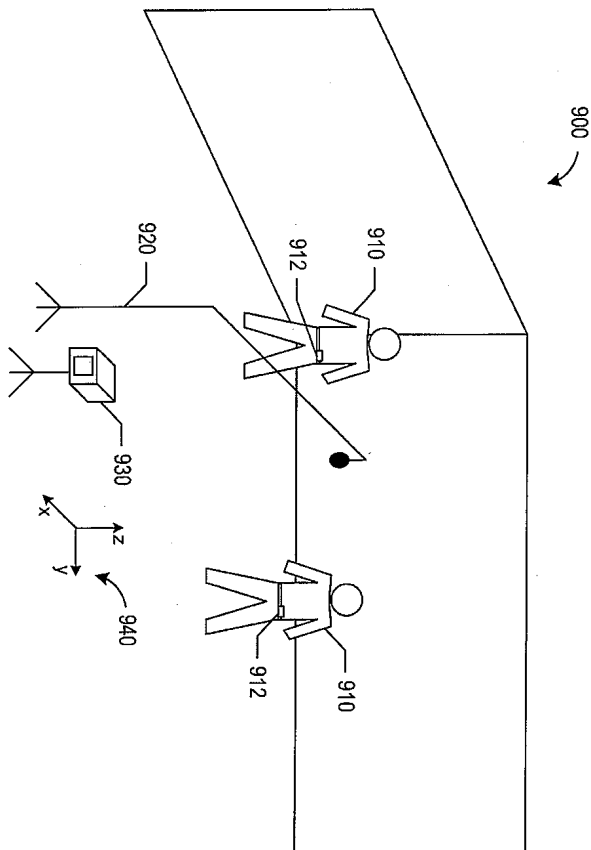
도면7



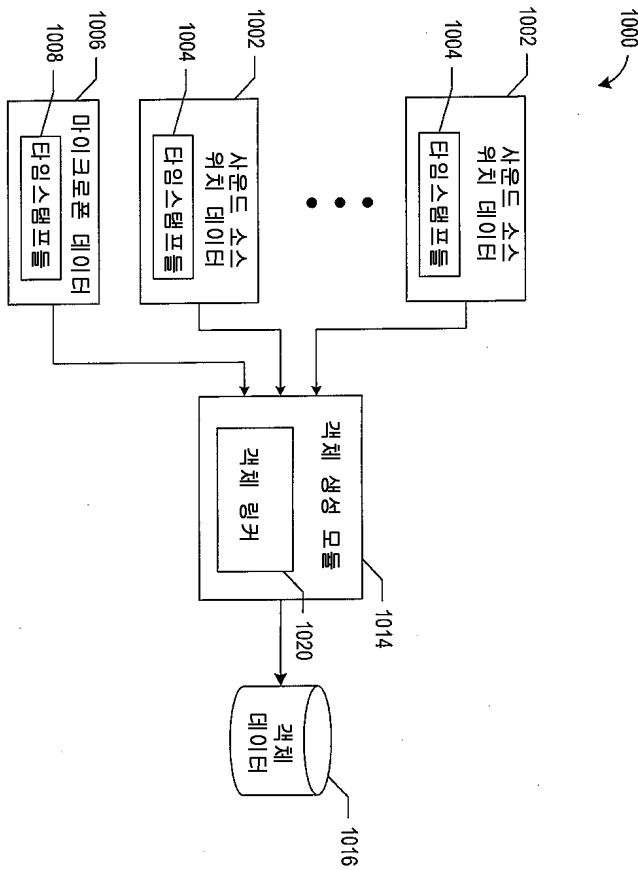
도면8



도면9



도면10



도면11

