



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0084113
(43) 공개일자 2022년06월21일

(51) 국제특허분류(Int. Cl.)
G10L 19/008 (2014.01) H04N 21/435 (2011.01)
H04N 21/439 (2011.01) H04S 7/00 (2006.01)
(52) CPC특허분류
G10L 19/008 (2020.08)
H04N 21/435 (2013.01)
(21) 출원번호 10-2022-7016218
(22) 출원일자(국제) 2020년10월08일
심사청구일자 없음
(85) 번역문제출일자 2022년05월13일
(86) 국제출원번호 PCT/EP2020/078297
(87) 국제공개번호 WO 2021/074007
국제공개일자 2021년04월22일
(30) 우선권주장
19202935.3 2019년10월14일
유럽특허청(EPO)(EP)

(71) 출원인
코닌클리케 필립스 엔.브이.
네덜란드 아인트호벤 5656 에이지 하이 테크 캠퍼스 52
(72) 발명자
딜렌 파울루스 헨리쿠스 안토니우스
네덜란드 아인트호벤 5656 에이이 하이 테크 캠퍼스 5 필립스 인터내셔널 비.브이. 인텔렉추얼 프라퍼티 앤드 스탠다즈 내
데 본트 프란시스쿠스 마리누스 요제푸스
네덜란드 아인트호벤 5656 에이이 하이 테크 캠퍼스 5 필립스 인터내셔널 비.브이. 인텔렉추얼 프라퍼티 앤드 스탠다즈 내
코펜스 제윈 게라두스 헨리쿠스
네덜란드 아인트호벤 5656 에이이 하이 테크 캠퍼스 5 필립스 인터내셔널 비.브이. 인텔렉추얼 프라퍼티 앤드 스탠다즈 내
(74) 대리인
장훈

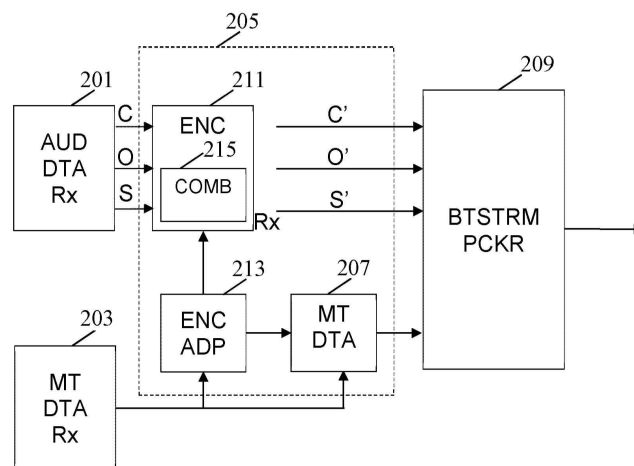
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 오디오 인코딩을 위한 장치 및 방법

(57) 요약

오디오 인코딩 장치는 오디오 장면을 나타내는 오디오 아이тем들을 수신하는 오디오 수신기(201)를 포함하고, 메타데이터 수신기(203)는 오디오 아이тем들의 렌더링에 대한 프레젠테이션 제약들을 설명하는 오디오 아이тем들에 대한 입력 프레젠테이션 메타데이터를 수신한다. 프레젠테이션 제약들은 오디오 아이тем들을 렌더링할 때 적용될 (뒷면에 계속)

대표도 - 도2



수 있는 렌더링 파라미터를 제약한다. 오디오 인코더(205)는 복수의 오디오 아이템들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하며, 이때 인코딩은 입력 프레젠테이션 메타데이터에 응답하여 적응된다. 메타데이터 회로(207)는 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성한다. 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이템들을 렌더링할 때 렌더링의 적응 가능 파라미터가 적응될 수 있는 정도를 제약하는 인코딩된 오디오 아이템들에 대한 데이터를 포함한다. 출력(209)은 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성한다.

(52) CPC특허분류

H04N 21/4394 (2013.01)

H04S 7/00 (2013.01)

명세서

청구범위

청구항 1

오디오 인코딩 장치로서,

오디오 장면(audio scene)을 나타내는 복수의 오디오 아이тем들을 수신하기 위한 오디오 수신기(201);

상기 복수의 오디오 아이тем들에 대한 입력 프레젠테이션 메타데이터를 수신하기 위한 메타데이터 수신기(203)로서, 상기 입력 프레젠테이션 메타데이터는 상기 복수의 오디오 아이тем들의 렌더링에 대한 프레젠테이션 제약들을 설명하고, 상기 프레젠테이션 제약들은 상기 복수의 오디오 아이тем들을 렌더링할 때 적용될 수 있는 렌더링 파라미터를 제약하는, 상기 메타데이터 수신기(203);

상기 복수의 오디오 아이тем들을 인코딩함으로써 상기 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하기 위한 오디오 인코더(205)로서, 상기 인코딩은 상기 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 오디오 인코더(205);

상기 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하기 위한 메타데이터 회로(207)로서, 상기 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이тем들을 렌더링할 때 렌더링의 적응 가능 파라미터가 적용될 수 있는 정도를 제약하는 상기 인코딩된 오디오 아이тем들에 대한 데이터를 포함하는, 상기 메타데이터 회로(207); 및

상기 인코딩된 오디오 데이터 및 상기 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하기 위한 출력 회로(209)를 포함하는, 오디오 인코딩 장치.

청구항 2

제1항에 있어서, 상기 오디오 인코더(205)는 상기 복수의 오디오 아이тем들 중 적어도 제1 오디오 아이тем과 제2 오디오 아이тем을 상기 제1 오디오 아이тем에 대한 입력 프레젠테이션 메타데이터 및 상기 제2 오디오 아이тем에 대한 입력 프레젠테이션 메타데이터에 응답하여 결합함으로써 결합된 오디오 아이тем을 생성하기 위한 결합기(215)를 포함하고, 상기 오디오 인코더(205)는 상기 결합된 오디오 아이тем을 인코딩함으로써 상기 제1 및 제2 오디오 아이тем에 대한 결합된 오디오 인코딩 데이터를 생성하도록, 그리고 상기 결합된 오디오 인코딩 데이터를 상기 인코딩된 오디오 데이터에 포함시키도록 배열되는, 오디오 인코딩 장치.

청구항 3

제2항에 있어서, 상기 결합기(215)는 상기 제1 오디오 아이тем 및 상기 제2 오디오 아이тем에 대한 상기 입력 프레젠테이션 메타데이터에 응답하여 상기 복수의 오디오 아이тем들로부터 상기 제1 오디오 아이тем 및 상기 제2 오디오 아이тем을 선택하도록 배열되는, 오디오 인코딩 장치.

청구항 4

제2항 또는 제3항에 있어서, 상기 결합기(215)는 적어도 몇몇의 상기 제1 오디오 아이тем에 대한 입력 프레젠테이션 메타데이터와 상기 제2 오디오 아이тем에 대한 입력 프레젠테이션 메타데이터가 유사성 기준을 충족시킨다는 결정에 응답하여 상기 제1 오디오 아이тем 및 상기 제2 오디오 아이тем을 선택하도록 배열되는, 오디오 인코딩 장치.

청구항 5

제2항 내지 제4항 중 어느 한 항에 있어서, 상기 제1 오디오 아이тем에 대한 상기 입력 프레젠테이션 메타데이터 및 상기 제2 오디오 아이тем에 대한 상기 입력 프레젠테이션 메타데이터는 이득 제약 및 위치 제약 중 적어도 하나를 포함하는, 오디오 인코딩 장치.

청구항 6

제2항 내지 제5항 중 어느 한 항에 있어서, 상기 오디오 인코더(205)는 상기 제1 오디오 아이템에 대한 상기 입력 프레젠테이션 메타데이터 및 상기 제2 오디오 아이템에 대한 상기 입력 프레젠테이션 메타데이터에 응답하여 상기 결합된 오디오 아이템에 대한 결합된 프레젠테이션 메타데이터를 생성하도록, 그리고 상기 결합된 프레젠테이션 메타데이터를 상기 출력 프레젠테이션 메타데이터에 포함시키도록 추가로 배열되는, 오디오 인코딩 장치.

청구항 7

제6항에 있어서, 상기 오디오 인코더(205)는 상기 결합된 오디오 아이템에 대한 프레젠테이션 파라미터에 대한 제약을 반영하기 위해 적어도 몇몇의 결합된 프레젠테이션 메타데이터를 생성하도록 배열되고, 상기 제약은 상기 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 상기 제1 오디오 아이템에 대한 제약 및 상기 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 상기 제2 오디오 아이템에 대한 제약 둘 모두를 충족시키는 제약으로서 결정되는, 오디오 인코딩 장치.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 오디오 인코더(205)는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.

청구항 9

제8항에 있어서, 상기 오디오 인코더(205)는 상기 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 상기 제2 오디오 아이템으로부터의 상기 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록, 그리고 상기 마스킹 효과에 응답하여 상기 제1 오디오 아이템의 상기 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.

청구항 10

제9항에 있어서, 상기 오디오 인코더(205)는 상기 제2 오디오 아이템에 대한 상기 입력 프레젠테이션 메타데이터에 의해 표시된 상기 제2 오디오 아이템에 대한 이득 제약 및 위치 제약 중 적어도 하나에 응답하여 상기 제2 오디오 아이템으로부터의 상기 제1 오디오 아이템에 대한 상기 마스킹 효과를 추정하도록 배열되는, 오디오 인코딩 장치.

청구항 11

제8항 내지 제10항 중 어느 한 항에 있어서, 상기 오디오 인코더(205)는 상기 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 상기 제1 오디오 아이템의 상기 압축을 적응시키도록 추가로 배열되는, 오디오 인코딩 장치.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 상기 입력 프레젠테이션 메타데이터는 적어도 몇몇의 오디오 아이템들에 대한 우선순위 데이터를 포함하고, 상기 인코더는 상기 입력 프레젠테이션 메타데이터 내의 제1 오디오 아이템에 대한 우선순위 표시에 응답하여 상기 제1 오디오 아이템에 대한 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, 상기 오디오 인코더(205)는 상기 인코딩이 상기 입력 프레젠테이션 메타데이터에 응답하여 어떻게 적응되는지를 나타내는 인코딩 적응 데이터를 생성하도록, 그리고 상기 인코딩 적응 데이터를 상기 인코딩된 오디오 데이터 스트림에 포함시키도록 배열되는, 오디오 인코딩 장치.

청구항 14

오디오를 인코딩하는 방법으로서,

오디오 장면을 나타내는 복수의 오디오 아이템들을 수신하는 단계;

상기 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계로서, 상기 입력 프레젠테

이션 메타데이터는 상기 복수의 오디오 아이템들의 렌더링에 대한 프레젠테이션 제약들을 설명하고, 상기 프레젠테이션 제약들은 상기 오디오 아이템들을 렌더링할 때 적용될 수 있는 렌더링 파라미터를 제약하는, 상기 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계;

상기 복수의 오디오 아이템들을 인코딩함으로써 상기 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계로서, 상기 인코딩은 상기 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 복수의 오디오 아이템들을 인코딩함으로써 상기 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계;

상기 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하는 단계로서, 상기 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이템들을 렌더링할 때 렌더링의 적응 가능 파라미터가 적용될 수 있는 정도를 제약하는 상기 인코딩된 오디오 아이템들에 대한 데이터를 포함하는, 상기 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하는 단계; 및

상기 인코딩된 오디오 데이터 및 상기 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하는 단계를 포함하는, 오디오를 인코딩하는 방법.

청구항 15

컴퓨터 프로그램 제품으로서, 상기 프로그램이 컴퓨터 상에서 실행될 때 제14항의 단계들 모두를 수행하도록 적응된 컴퓨터 프로그램 코드 수단을 포함하는, 컴퓨터 프로그램 제품.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 인코딩을 위한, 그리고 특히, 그러나 비배타적으로, 가상 현실 응용들과 같은 동적 응용들을 위한 오디오의 인코딩을 위한 장치 및 방법에 관한 것이다.

배경 기술

[0002] 오디오 및 비디오 응용들의 다양성 및 범위는 최근 수년간 상당히 증가했으며, 오디오, 이미지들 및 비디오를 이용하고 소비하는 새로운 서비스들 및 방식들이 계속해서 개발 및 도입되고 있다.

[0003] 예를 들어, 점점 더 인기가 있는 하나의 서비스는 관찰자(viewer)가 렌더링의 파라미터들을 변경하기 위해 시스템과 능동적으로 그리고 동적으로 상호작용할 수 있는 방식으로 오디오 및 이미지들을 제공하는 것이다. 많은 응용들에서 매우 매력적인 특징은 유효 관찰/청취 위치를 변경하는 능력이다. 그러한 특징은 특히 가상 현실 경험이 사용자에게 제공될 수 있게 할 수 있다.

[0004] 추세는 장면(scene)의 렌더링 측 적응을 가능하게 하는 증가하는 유연성을 제공하는 것을 향한다. 오디오 장면(audio scene)의 렌더링을 위한 증가된 렌더링 측 유연성을 제공하기 위해, 오디오 장면이 상이한 오디오 아이템들의 결합들에 의해 표현될 수 있는 다수의 오디오 인코딩 및 분배 접근법들이 제안되었다. 예를 들어, 오디오 아이템들은 특정 스피커들 등과 같은, 별개의 사운드 소스들을 표현할 수 있다. 몇몇 접근법들에서, 모든 오디오 아이템들은 동일한 유형의 것이지만, 다수의 상이한 오디오 유형들이 동시에 사용 및 지원될 수 있게 하는 점점 더 많은 시스템들이 개발된다. 예를 들어, 일부 오디오 아이템들은 오디오 채널들일 수 있고, 다른 것들은 별개의 오디오 객체들일 수 있고, 또 다른 것은 앰비소닉(Ambisonic) 오디오 아이템들과 같은 장면 기반일 수 있고, 등등이다. 많은 시스템들에서, 메타데이터가 오디오 아이템들을 표현하는 오디오 데이터와 함께 제공될 수 있다. 그러한 메타데이터는 예를 들어 오디오 아이템의 오디오 소스에 대한 장면 내의 공칭 위치를 표시할 수 있다.

[0005] 그러한 접근법들은 고도의 클라이언트/렌더링 측 맞춤화 및 적응을 가능하게 할 수 있다. 예를 들어, 오디오 장면은 오디오 장면 내의 청취자 가상 위치의 변경들에, 또는 개별 청취자의 특정 선호들에 로컬에서 적응될 수 있다.

[0006] 특정 예로서, 3GPP 컨소시엄이 이른바 IVAS(Immersive Voice and Audio Services) 코덱을 현재 개발하고 있다. 이러한 코덱은 채널, 객체, 또는 장면(특히, 앰비소닉스) 기반 구성들과 같은, 다양한 구성들에서 오디오 콘텐츠들을 코딩할 수 있을 것이다. 코딩의 목적은 최소량의 데이터로 오디오 정보를 전달하는 것이다.

[0007] IVAS 코덱은 더욱이 다양한 오디오 스트림들을 수신단에서의 재생에 적합한 형태로 변환하는 렌더러를 포함할

것이다. 예를 들어, 오디오가 공지된 라우드스피커 구성에 매핑될 수 있거나, 오디오가 헤드폰들을 통한 재생을 위해 바이노럴 포맷(binaural format)으로 렌더링될 수 있다.

[0008] 3GPP IVAS 코덱 범위에서, 잠재적 사용 사례들을 수집하기 위한 연구가 계속 진행 중이다. 이것들을 위해, 코덱은 렌더링을 변조하기 위한 쌍방향성을 제공해야 하는 것으로 예상된다. 예를 들어, 헤드폰 오디오는 머리 위치 및 병진과 독립적으로 렌더링되어야 할 수 있으며, 이는 그것이 머리 움직임에 대해 보상되어야 한다는 것을 의미한다. 다른 예로서, 사용자는 가상 회의에서 참가자들의 오디오를 운반하는 객체들을 (재)배치하는 것과 같이, 오디오 아이тем들을 공간적으로 배치할 수 있게 될 수 있다.

[0009] 렌더러는 3GPP IVAS 코덱 작업 아이тем의 일부로 간주되고, IVAS 코덱의 내부에 있는 것으로 여겨질 수 있다. 그러나, 코덱이 또한 통과 모드(pass-through mode)를 포함하는 것이 제안되었다. 이러한 모드는 오디오 아이тем들이, 그들이 인코더 입력에서 입력되었던 것과 동일한 구성(들)으로(즉, 1:1 대응하는 채널, 객체 및 장면 기반 오디오 아이тем들로서) 디코더 출력에서 표현될 수 있게 할 것이다. 전용 외부 렌더링 인터페이스를 통해, 외부 렌더러가 이러한 아이тем들에 액세스할 수 있고 내부 IVAS 렌더러에 대한 대안적인 렌더링을 실현할 수 있다.

[0010] 그러한 접근법은 추가적인 유연성을 제공하고 수신단에서의 맞춤화 및 적응을 위한 범위를 증가시킬 수 있다. 그러나, 이 접근법은 또한 연관된 불리한 점들을 가질 수 있다. 예를 들어, 유연성 대 오디오 품질 및 복잡도 사이에 트레이드-오프가 존재한다. 자유를 제한하여서, 콘텐츠 제공자가 클라이언트 측에서 렌더링의 얼마간의 제어를 유지할 수 있게 하는 것이 전형적으로 유용할 수 있다. 이것은 렌더링을 돕고 더 현실적인 렌더링된 오디오 장면을 야기할 수 있을 뿐만 아니라, 콘텐츠 제공자가 사용자에게 제공되는 경험에 대한 얼마간의 제어를 유지할 수 있게 할 수 있다. 예를 들어, 그것은 렌더러가 비현실적이고 콘텐츠 및 콘텐츠 제공자에게 나쁜 영향을 미칠 수 있는 오디오 장면을 생성하는 것을 방지할 수 있다.

[0011] 인코딩된 오디오 아이тем들이 어떻게 렌더러가 오디오 아이тем들을 렌더링하도록 허용되는지를 제약하는 메타데이터에 의해 보충될 수 있는 것이 예상된다. 이것은 많은 상황들에서 상이한 요건들 사이의 개선된 트레이드-오프를 허용할 수 있다. 그러나, 그것은 모든 상황들에서 최적인 아닐 수 있고 예를 들어 증가된 데이터 레이트를 요구할 수 있으며, 렌더링된 오디오 장면에 대한 감소된 유연성 및/또는 품질을 야기할 수 있다.

[0012] 따라서, 개선된 접근법이 바람직할 것이다. 특히, 개선된 동작, 증가된 유연성, 용이한 구현, 용이한 동작, 감소된 비용, 감소된 복잡도, 감소된 데이터 레이트, 개선된 인지된 오디오 품질, 개선된 렌더링 제어, 개선된 트레이드-오프, 및/또는 개선된 성능을 허용하는 접근법이 유리할 것이다.

발명의 내용

[0013] 따라서, 본 발명은 전술한 불리한 점들 중 하나 이상을 단독으로 또는 임의의 조합으로 바람직하게 완화, 경감 또는 제거하고자 한다.

[0014] 본 발명의 태양에 따르면, 오디오 인코딩 장치로서, 오디오 장면을 나타내는 복수의 오디오 아이тем들을 수신하기 위한 오디오 수신기; 복수의 오디오 아이тем들에 대한 입력 프레젠테이션 메타데이터를 수신하기 위한 메타데이터 수신기로서, 입력 프레젠테이션 메타데이터는 복수의 오디오 아이тем들의 렌더링에 대한 프레젠테이션 제약들을 설명하고, 프레젠테이션 제약들은 오디오 아이тем들을 렌더링할 때 적용될 수 있는 렌더링 파라미터를 제약하는, 상기 메타데이터 수신기; 복수의 오디오 아이тем들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하기 위한 오디오 인코더로서, 인코딩은 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 오디오 인코더; 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하기 위한 메타데이터 회로로서, 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이тем들을 렌더링할 때 렌더링의 적응 가능 파라미터가 적용될 수 있는 정도를 제약하는 인코딩된 오디오 아이тем들에 대한 데이터를 포함하는, 상기 메타데이터 회로; 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하기 위한 출력 회로를 포함하는, 오디오 인코딩 장치가 제공된다.

[0015] 본 발명은 많은 시나리오들에서 개선된 그리고/또는 더 유연한 인코딩을 제공할 수 있다. 이 접근법은 많은 실시예들에서 개선된 품질 대 비트 레이트 비율을 제공하는 인코딩된 오디오 데이터 스트림이 생성될 수 있게 할 수 있다. 인코딩된 오디오 데이터 스트림은 얼마간의 렌더링의 유연성을 허용하면서 또한 소스/인코딩 측으로부터의 렌더링의 얼마간의 제어를 허용하도록 생성될 수 있다.

[0016] 오디오 아이тем에 대한 프레젠테이션 메타데이터는, 예를 들어 렌더링 위치, 이득 레벨, 신호 레벨, 공간 분배,

또는 반향 특성을 제약하는 것을 포함하여, 오디오 아이템의 렌더링에 대한 공간 파라미터 및 볼륨 파라미터 중 적어도 하나를 제약할 수 있다.

- [0017] 오디오 인코더는 입력 프레젠테이션 메타데이터에 기초하여, 그리고 구체적으로 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 기초하여 오디오 아이템의 인코딩을 적응시키도록 배열될 수 있다. 적응은 오디오 아이템의 인코딩에 대한 비트/데이터 (레이트) 압축을 적응시킬 수 있다. 오디오 아이템을 인코딩하는 것으로부터 기인하는 비트레이트는 입력 프레젠테이션 메타데이터에 기초하여 적응될 수 있다.
- [0018] 입력 프레젠테이션 메타데이터는 수신된 복수의 오디오 아이템들에 대한 프레젠테이션/렌더링 제약들을 설명할 수 있다. 인코딩된 오디오 데이터는 복수의 인코딩된 오디오 아이템들에 대한 오디오 데이터를 포함할 수 있다. 복수의 인코딩된 오디오 아이템들은 수신된 복수의 오디오 아이템들의 인코딩에 의해 생성될 수 있다. 복수의 인코딩된 오디오 아이템들의 렌더링에 대한 프레젠테이션/렌더링 제약들을 설명하는 출력 프레젠테이션 메타데이터.
- [0019] 프레젠테이션 제약은 렌더링 제약일 수 있고, 오디오 아이템에 대한 렌더링 파라미터를 제약할 수 있다. 렌더링 파라미터는 렌더링 프로세스의 파라미터 및/또는 렌더링된 신호의 특성일 수 있다.
- [0020] 출력 프레젠테이션 메타데이터는 구체적으로 프레젠테이션/렌더링의 하나 이상의 적응 가능한/가변적인 태양들/특성들/파라미터들이 인코딩된 오디오 아이템들을 렌더링할 때 적응될 수도/적용될 수 있는 정도를 제약하는 오디오 인코더에 의해 생성된 인코딩된 오디오 아이템들과 연관된/링크된/그들에 대해 제공된 임의의 데이터일 수 있다.
- [0021] 출력 프레젠테이션 메타데이터, 그리고 구체적으로 렌더링의 적응 가능한 파라미터가 인코딩된 오디오 아이템들을 렌더링할 때 적용될 수 있는 정도를 제약하는 인코딩된 오디오 아이템들에 대한 데이터는, 복수의 오디오 아이템들을 렌더링할 때 적용될 수 있는 렌더링 파라미터를 제약하는 프레젠테이션 제약들에 응답하여 메타데이터 회로에 의해 생성될 수 있다.
- [0022] 오디오 인코더는 (복수의 오디오 아이템들을 인코딩함으로써) 복수의 인코딩된 오디오 아이템들을 포함하도록 인코딩된 오디오 데이터를 생성할 수 있다.
- [0023] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 복수의 오디오 아이템들 중 적어도 제1 오디오 아이템과 제2 오디오 아이템을 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 결합함으로써 결합된 오디오 아이템을 생성하기 위한 결합기를 포함하고, 오디오 인코더는 결합된 오디오 아이템을 인코딩함으로써 제1 및 제2 오디오 아이템에 대한 결합된 오디오 인코딩 데이터를 생성하도록, 그리고 결합된 오디오 인코딩 데이터를 인코딩된 오디오 데이터에 포함시키도록 배열된다.
- [0024] 이것은 많은 실시예들에서 특히 효율적인 인코딩 및/또는 유연성을 제공할 수 있다. 그것은 특히 많은 실시예들에서 감소된 지각적 저하를 갖는 효율적인 비트레이트 압축을 제공할 수 있다.
- [0025] 본 발명의 선택적인 특징에 따르면, 결합기는 제1 오디오 아이템 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 복수의 오디오 아이템들로부터 제1 오디오 아이템 및 제2 오디오 아이템을 선택하도록 배열된다.
- [0026] 이것은 많은 실시예들에서 특히 효율적인 인코딩 및/또는 유연성을 제공할 수 있다.
- [0027] 본 발명의 선택적인 특징에 따르면, 결합기는 적어도 몇몇의 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터와 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터가 유사성 기준을 충족시킨다는 결정에 응답하여 제1 오디오 아이템 및 제2 오디오 아이템을 선택하도록 배열된다.
- [0028] 이것은 많은 실시예들에서 특히 효율적인 인코딩 및/또는 유연성을 제공할 수 있다. 유사성 기준은 프레젠테이션 메타데이터에 의해 제약된 렌더링 파라미터에 대한 렌더링 제약들이 유사성 기준을 충족시키는 요건을 포함할 수 있다.
- [0029] 본 발명의 선택적인 특징에 따르면, 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터는 이득 제약 및 위치 제약 중 적어도 하나를 포함한다.
- [0030] 이것은 많은 실시예들에서 특히 효율적인 동작을 제공할 수 있다.

- [0031] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 결합된 오디오 아이템에 대한 결합된 프레젠테이션 메타데이터를 생성하도록, 그리고 결합된 프레젠테이션 메타데이터를 출력 프레젠테이션 메타데이터에 포함시키도록 추가로 배열된다.
- [0032] 이것은 많은 실시예들에서 개선된 동작을 제공할 수 있고, 특히 많은 실시예들에서 인코더가 결합된 오디오 아이템들과 인코딩된 입력 오디오 아이템들을 동일한 방식으로, 그리고 실제로 개개의 오디오 아이템이 결합된 오디오 아이템인지 여부에 대한 지식 없이 취급할 수 있게 할 수 있다.
- [0033] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 결합된 오디오 아이템에 대한 프레젠테이션 파라미터에 대한 제약을 반영하기 위해 적어도 몇몇의 결합된 프레젠테이션 메타데이터를 생성하도록 배열되고, 제약은 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제1 오디오 아이템에 대한 제약 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제2 오디오 아이템에 대한 제약 둘 모두를 충족시키는 제약으로서 결정된다.
- [0034] 이것은 많은 시나리오들 및 응용들에서 개선된 성능을 제공할 수 있다.
- [0035] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 배열된다.
- [0036] 이 접근법은 전형적으로 오디오 아이템의 개선된 압축 및 인코딩을 가능하게 할 수 있다. 압축은 비트 레이트 감소될 수 있고 증가하는 압축은 인코딩된 오디오 아이템의 감소된 데이터 레이트를 야기할 수 있다. 압축은 비트 레이트 감소/압축될 수 있다. 오디오 인코딩은 하나 이상의 입력 오디오 아이템을 표현하는 인코딩된 오디오 아이템이 하나 이상의 입력 오디오 아이템보다 더 적은 비트에 의해 표현되도록 하는 것일 수 있다.
- [0037] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록, 그리고 마스킹 효과에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 배열된다.
- [0038] 이것은 많은 실시예들에서 특히 효율적인 동작 및 개선된 성능을 제공할 수 있다.
- [0039] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제2 오디오 아이템에 대한 이득 제약 및 위치 제약 중 적어도 하나에 응답하여 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록 배열된다.
- [0040] 이것은 많은 실시예들에서 특히 효율적인 동작 및 개선된 성능을 제공할 수 있다.
- [0041] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 추가로 배열된다.
- [0042] 이것은 많은 실시예들에서 특히 유리한 동작 및/또는 성능을 제공할 수 있다.
- [0043] 본 발명의 선택적인 특징에 따르면, 입력 프레젠테이션 메타데이터는 적어도 몇몇의 오디오 아이템들에 대한 우선순위 데이터를 포함하고, 인코더는 입력 프레젠테이션 메타데이터 내의 제1 오디오 아이템에 대한 우선순위 표시에 응답하여 제1 오디오 아이템에 대한 압축을 적응시키도록 배열된다.
- [0044] 이것은 많은 실시예들에서 특히 유리한 동작 및/또는 성능을 제공할 수 있다.
- [0045] 본 발명의 선택적인 특징에 따르면, 오디오 인코더는 인코딩이 입력 프레젠테이션 메타데이터에 응답하여 어떻게 적응되는지를 나타내는 인코딩 적응 데이터를 생성하도록, 그리고 인코딩 적응 데이터를 인코딩된 오디오 데이터 스트림에 포함시키도록 배열된다.
- [0046] 이것은 많은 실시예들에서 특히 유리한 동작 및/또는 성능을 제공할 수 있다. 그것은 특히 디코더에 의한 개선된 적응이 인코딩 프로세스와 매칭될 수 있게 할 수 있다.
- [0047] 본 발명의 태양에 따르면, 오디오를 인코딩하는 방법으로서, 오디오 장면을 나타내는 복수의 오디오 아이템들을 수신하는 단계; 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계로서, 입력 프레젠테이션 메타데이터는 복수의 오디오 아이템들의 렌더링에 대한 프레젠테이션 제약들을 설명하고, 프레젠테이션 제약들은 오디오 아이템들을 렌더링할 때 적응될 수 있는 렌더링 파라미터를 제약하는, 상기 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계; 복수의 오디오 아이템들을 인코딩함으로써 오

디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계로서, 인코딩은 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 복수의 오디오 아이템들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계; 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하는 단계로서, 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이템들을 렌더링할 때 렌더링의 적응 가능 파라미터가 적응될 수 있는 정도를 제약하는 인코딩된 오디오 아이템들에 대한 데이터를 포함하는, 상기 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하는 단계; 및 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하는 단계를 포함하는, 오디오를 인코딩하는 방법이 제공된다.

[0048] 본 발명의 이들 및 다른 태양들, 특징들 및 이점들이 이하에 설명되는 실시예(들)로부터 명백할 것이며 그것을 참조하여 설명될 것이다.

도면의 간단한 설명

[0049] 본 발명의 실시예들이 도면을 참조하여 단지 예로서 설명될 것이다.

도 1은 본 발명의 몇몇 실시예들에 따른 오디오 분배 시스템의 요소들의 예를 예시한다.

도 2는 본 발명의 몇몇 실시예들에 따른 오디오 인코딩 장치의 요소들의 예를 예시한다.

도 3은 본 발명의 몇몇 실시예들에 따른 오디오 디코딩 장치의 요소들의 예를 예시한다.

발명을 실시하기 위한 구체적인 내용

[0050] 이하의 설명은 3GPP IVAS(Immersive Voice and Audio Services) 코덱과 호환 가능할 수 있는 오디오 인코딩 및 디코딩 시스템에 초점을 맞출 것이지만, 설명된 원리들 및 개념들이 많은 다른 응용들 및 실시예들에서 사용될 수 있다는 것이 인식될 것이다.

[0051] 도 1은 오디오 인코딩 시스템의 예를 예시한다. 시스템에서, 오디오 소스(101)는 오디오 데이터를 오디오 인코더 유닛(103)에 제공한다. 오디오 데이터는 오디오 장면의 오디오를 나타내는 복수의 오디오 아이템들에 대한 오디오 데이터를 포함한다. 오디오 아이템들은 구체적으로 다음을 포함한 상이한 유형들로서 제공될 수 있다.

[0052] 채널 기반 오디오 아이템들: 그러한 오디오 아이템들에 대해, 1D(모노럴(monaural)), 2D 또는 3D 공간 오디오 콘텐츠는 전형적으로, 청취자에 대한 미리 결정된 위치들에서 라우드스피커들을 통해 제시되도록 의도되는, 개별 신호들로서 표현된다. 잘 알려진 라우드스피커 셋업들은 예를 들어 2-채널 스테레오('2.0'으로도 알려짐), 또는 청취자 주위의 5개의 채널 + 저주파 효과 채널('5.1'로도 지칭됨)이다. 또한 바이노럴 오디오가 전형적으로 (통상적으로 헤드폰들을 통해) 청취자의 각자의 귀에 직접 제시되도록 의도된 2개의 오디오 신호 채널로 이루어진 채널 기반 오디오인 것으로 간주된다.

[0053] 객체 기반 오디오 아이템들: 그러한 오디오 아이템들에 대해, 개개의 오디오 신호들이 전형적으로 별개의 사운드 소스들을 표현하는 데 사용된다. 이러한 사운드 소스들은 종종 실제 객체들 또는 사람들, 예를 들어 원격 회의의 참가자들과 관련된다. 신호들은 전형적으로 모노이지만, 다른 표현들이 또한 사용될 수 있다. 객체 기반 오디오 신호들은 종종 객체 오디오의 범위(공간 확산), 지향성 또는 확산성과 같은, 추가 특성들을 설명하는 메타데이터를 동반한다.

[0054] 장면 기반 오디오 아이템들: 그러한 오디오 아이템들에 대해, 오리지널 2D 또는 3D 공간 오디오 장면은 전형적으로 소정의 구면 조화 함수와 관련된 다수의 오디오 신호로서 표현된다. 이러한 장면 기반 오디오 신호들을 결합함으로써, 제시 가능한 오디오 신호들이 임의적인 2D 또는 3D 위치들에서, 예를 들어 오디오 재생 셋업 내의 실제 라우드스피커의 위치에서 구성될 수 있다. 장면 기반 오디오의 예시적인 구현은 앰비소닉스이다. 장면-기반 오디오는 특정 라우드스피커 레이아웃들과는 무관한 라이브-캡처된 사운드 장면 및 예술적으로 생성된 사운드 장면 둘 모두의 전체론적 설명들을 생성하기 위해 "고차 앰비소닉스"(HOA)로 지칭되는 사운드-필드 기술을 사용한다.

[0055] 오디오 데이터에 더하여, 오디오 소스는 오디오 아이템들에 대한 프레젠테이션 메타데이터를 제공할 수 있다. 프레젠테이션 메타데이터는 오디오 장면의 렌더링에 대한 프레젠테이션 제약들을 설명할 수 있고, 이에 따라 복수의 오디오 아이템들에 대한 프레젠테이션/렌더링 제약들을 제공할 수 있다.

[0056] 프레젠테이션 메타데이터는 어떻게 오디오 아이템들의 렌더링이 렌더러에 의해 수행되어야 하는지에 있어서의

제약을 설명할 수 있다. 프레젠테이션 메타데이터는 하나 이상의 렌더링 파라미터/특성에 대한 제약을 정의할 수 있다. 파라미터/특성은 구체적으로 오디오 아이템의 렌더링의 지각 특성에 영향을 미치는 것일 수 있다. 제약은 장면 내의 오디오 아이템의 공간 인지 및/또는 (상대) 신호 레벨에 영향을 미치는 제약일 수 있다. 프레젠테이션 메타데이터는 구체적으로 하나 이상의 오디오 아이템에 대한 공간 및/또는 이득/신호 레벨 파라미터를 제약할 수 있다. 메타데이터는 예를 들어 각각의 오디오 아이템에 대한 위치 및/또는 이득에 대한 제약일 수 있다.

- [0057] 메타데이터는 예를 들어 하나 이상의 오디오 아이템의 하나 이상의 파라미터에 대한 허용 가능한 값들의 범위 또는 세트를 설명할 수 있다. 오디오 아이템(들)의 렌더링은 제약 내에서 자유롭게 수행될 수 있는데, 즉 렌더링은 제약된 파라미터들이 표시된 허용 가능한 값들 중 임의의 것을 갖도록 하는 것일 수 있지만, 제약된 파라미터가 이러한 값을 갖지 않도록 하는 것이 아닐 수 있다.
- [0058] 예로서, 프레젠테이션 메타데이터는 오디오 아이템들 중 하나 이상에 대해 영역 및/또는 (상대) 이득 범위를 설명할 수 있다. 그렇다면 오디오 아이템은 영역 내의 인지된 위치와 함께 그리고/또는 이득 범위 내의 이득과 함께 렌더링되어야 한다.
- [0059] 이에 따라 프레젠테이션 메타데이터는 로컬 렌더링을 적응시키고 맞춤화하는 얼마간의 유연성을 여전히 허용하면서 렌더링을 제약할 수 있다.
- [0060] 프레젠테이션 메타데이터에 의해 제공될 수 있는 파라미터들 또는 특성들의 렌더링 제약들의 예들은 다음을 포함한다:
- [0061] 하나 이상의 오디오 아이템에 대한 위치 제약. 이것은 예를 들어 오디오 아이템이 렌더링되어야 하는 오디오 장면 내의 공간 영역 또는 볼륨을 정의할 수 있다.
- [0062] 하나 이상의 오디오 아이템에 대한 반향 제약. 이것은 예를 들어 최소 또는 최대 반향 시간을 정의할 수 있다. 제약은 예를 들어 오디오 아이템이 원하는 정도의 확산성으로 렌더링되는 것을 보장할 수 있다. 예를 들어, 일 반적인 주변 배경 사운드를 나타내는 오디오 아이템은 최소량의 반향으로 렌더링되도록 요구될 수 있는 반면, 메인 스피커를 나타내는 오디오 아이템은 주어진 반향 임계치 미만으로 렌더링되도록 요구될 수 있다.
- [0063] 이득 제약. 오디오 아이템의 렌더링은 렌더링 프로세스의 특정 선호들에 따라 더 크거나 더 조용하도록 렌더러에 의해 적응될 수 있다. 예를 들어, 주변 배경 사운드들에 대한 스피커에 대한 이득은 몇몇 경우들에서 청취자 선호들에 기초하여 증가되거나 감소될 수 있다. 그러나, 이득 제약은 이득이 얼마나 많이 수정될 수 있는지를 제약할 수 있고, 예를 들어 그에 의해 스피커가 항상 주변 잡음보다 더 크게 들릴 수 있는 것을 보장할 수 있다.
- [0064] 라우드니스(loudness) 제약. 오디오 아이템의 렌더링은 렌더링 프로세스의 특정 선호들에 따라 더 크거나 더 조용하도록 렌더러에 의해 적응될 수 있다. 예를 들어, 원격 회의 참가자들에 대한 이득은 몇몇 경우들에서 청취자의 선호들에 기초하여 증가되거나 감소될 수 있다. 그러나, 라우드니스 제약은 소정 참가자들의 인지된 라우드니스가 얼마나 많이 수정될 수 있는지를 제약할 수 있고, 예를 들어 그에 의해 예를 들어 회의 의장이 다른 스피커들 또는 배경 잡음의 존재 시에 항상 충분히 크게 들릴 수 있는 것을 보장할 수 있다.
- [0065] 동적 범위 제어 제약. 오디오 아이템의 동적 범위는 렌더러에 의해 더 크도록 적응될 수 있는데, 예를 들어 그것은 오디오가 청취자 위치에서 배경 잡음의 존재 시에 더 낮은 레벨의 기간들 동안에도 가청 상태로 유지되도록 감소될 수 있다. 예를 들어, 바이올린 사운드가 자동으로 낮은 레벨들에서 더 커지게 될 수 있다. 그러나, 동적 범위 제어 제약은 동적 범위가 얼마나 많이 감소될 수 있는지를 제약할 수 있고, 이에 따라 예를 들어 바이올린의 정상 역학의 충분히 자연스러운 인지를 보장할 수 있다.
- [0066] 복수의 오디오 아이템들의 렌더링에 대한 프레젠테이션 제약들을 설명하는 프레젠테이션 메타데이터는 구체적으로 (프레젠테이션 메타데이터가 제공되는) 오디오 아이템들을 렌더링할 때 적응될 수 있는 렌더링 파라미터 또는 특성에 대한 제약들을 제공하는 데이터일 수 있다. 렌더링 파라미터 또는 특성은 렌더링 동작의 파라미터/특성, 및/또는 생성된 렌더링된/제시된 신호 및/또는 오디오의 파라미터 또는 특성일 수 있다.
- [0067] 입력 프레젠테이션 메타데이터는 구체적으로 프레젠테이션/렌더링의 하나 이상의 적응 가능한/가변적인 태양들/특성들/파라미터들이 입력 오디오 아이템들을 렌더링할 때 적응될 수도/적용될 수 있는 정도를 제약하는 오디오 인코더(205)에 대한 입력 오디오 아이템들과 연관된/링크된/그들에 대해 제공된 임의의 데이터일 수 있다.
- [0068] 오디오 인코더 유닛(103)은 오디오 장면에 대한 인코딩된 오디오 데이터를 포함하는 인코딩된 오디오 데이터 스

트림을 생성하도록 배열된다. 인코딩된 오디오 데이터는 오디오 아이템들(즉, 오디오 아이템들을 나타내는 수신된 오디오 데이터)을 인코딩함으로써 생성된다. 또한, 오디오 인코더 유닛(103)은 인코딩된 오디오 아이템들에 대한 출력 프레젠테이션 메타데이터를 생성하고, 이러한 메타데이터를 인코딩된 오디오 데이터 스트림에 포함시킨다. 출력 프레젠테이션 메타데이터는 인코딩된 오디오 아이템들에 대한 렌더링 제약들을 설명한다.

[0069] 출력 프레젠테이션 메타데이터는 구체적으로 프레젠테이션/렌더링의 하나 이상의 적응 가능한/가변적인 태양들/특성들/파라미터들이 인코딩된 오디오 아이템들을 렌더링할 때 적용될 수도/적용될 수 있는 정도를 제약하는 오디오 인코더(205)에 의해 생성된 인코딩된 오디오 아이템들과 연관된/링크된/그들에 대해 제공된 임의의 데이터일 수 있다.

[0070] 출력 프레젠테이션 메타데이터, 그리고 구체적으로 렌더링의 적응 가능한 파라미터가 인코딩된 오디오 아이템들을 렌더링할 때 적용될 수 있는 정도를 제약하는 인코딩된 오디오 아이템들에 대한 데이터는, 복수의 (입력) 오디오 아이템들을 렌더링할 때 적용될 수 있는 렌더링 파라미터를 제약하는 (입력) 프레젠테이션 제약들에 응답하여 메타데이터 회로에 의해 생성될 수 있다.

[0071] 오디오 인코더 유닛(103)은 인코딩된 오디오 데이터 스트림을 공급받는 송신기(105)에 결합된다. 송신기(105)는 이 예에서 인코딩된 오디오 데이터 스트림을, 인코딩된 오디오 데이터 스트림에 기초하여 오디오 장면을 렌더링할 수 있는 하나 이상의 클라이언트에 송신/분배하도록 배열된다.

[0072] 이 예에서, 인코딩된 오디오 데이터 스트림은 구체적으로 인터넷될 수 있거나 인터넷을 포함할 수 있는 네트워크(107)를 통해 분배된다. 송신기(105)는 잠재적으로 많은 수의 클라이언트를 동시에 지원하도록 배열될 수 있고 오디오 데이터는 일반적으로 복수의 클라이언트에 분배될 수 있다.

[0073] 특정 예에서, 인코딩된 오디오 데이터 스트림은 하나 이상의 렌더링 디바이스(109)에 송신될 수 있다. 렌더링 디바이스(109)는 네트워크(107)로부터 인코딩된 오디오 데이터 스트림을 수신하는 수신기(111)를 포함할 수 있다.

[0074] 송신기(105)와 수신기(111)는 임의의 적합한 형태로 그리고 임의의 적합한 통신 프로토콜, 표준, 기술 및 기능을 사용하여 통신할 수 있다는 것이 인식될 것이다. 이 예에서, 송신기(105)와 수신기(111)는 적절한 네트워크 인터페이스 기능을 포함할 수 있지만, 다른 실시예들에서, 송신기(105) 및/또는 수신기(111)는 예를 들어 무선 통신 기능, 광섬유 통신 기능 등을 포함할 수 있다는 것이 인식될 것이다.

[0075] 수신기(111)는 수신된 인코딩된 오디오 데이터 스트림을 공급받는 디코더(113)에 결합된다. 디코더(113)는 인코딩된 오디오 데이터 스트림을 디코딩하여 오디오 아이템들을 재생성하도록 배열된다. 디코더(113)는 인코딩된 오디오 데이터 스트림으로부터 프레젠테이션 메타데이터를 추가로 디코딩할 수 있다.

[0076] 디코더(113)는 오디오 아이템들에 대한 디코딩된 오디오 데이터 및 프레젠테이션 메타데이터를 공급받는 렌더러(115)에 결합된다. 렌더러(115)는 수신된 프레젠테이션 메타데이터에 기초하여 오디오 아이템들을 렌더링함으로써 오디오 장면을 렌더링할 수 있다. 렌더러(115)에 의한 렌더링은 사용되는 특정 오디오 재생 시스템을 타겟으로 할 수 있다. 예를 들어, 5.1 서라운드 사운드 시스템에 대해, 개별 채널들에 대한 오디오 신호들이 생성될 수 있고, 헤드폰 시스템에 대해, 바이노럴 신호들이 예를 들어 HRTF 필터들을 사용하여 생성될 수 있고, 등등이다. 많은 상이한 가능한 오디오 렌더링 알고리즘들 및 기술들이 알려져 있고, 본 발명으로부터 벗어남이 없이 임의의 적합한 접근법이 사용될 수 있다는 것이 인식될 것이다.

[0077] 렌더러(115)는 구체적으로 결합된 재생이 청취자에 의해 인지될 때 오디오 장면의 인지를 제공하도록 재생을 위한 출력 오디오 신호들을 생성할 수 있다. 렌더러는 전형적으로 개별 오디오 아이템에 대한 특정 특성들에 따라 상이한 오디오 아이템들을 개별적으로 그리고 상이하게 처리하고, 이어서 각각의 출력 채널에 대한 결과적인 신호 성분들을 결합할 것이다. 예를 들어, 오디오 객체 오디오 아이템에 대해, 오디오 객체에 대응하는 오디오 소스에 대한 오디오 장면 내의 원하는 위치에 따라 각각의 출력 채널에 대해 신호 성분들이 생성될 수 있다. 오디오 채널 오디오 아이템은 예를 들어 대응하는 출력 재생 채널에 대한 신호 성분의 생성에 의해, 또는 예를 들어 (예를 들어, 적절한 경우 패닝(panning) 또는 업믹싱(upmixing) 기술들을 사용하여) 그것이 재생 채널들 중 하나에 정확하게 매핑되지 않는 경우 복수의 재생 채널들에 의해 렌더링될 수 있다.

[0078] 다수의 전형적으로 상이한 유형의 오디오 아이템들에 의한 오디오 장면의 표현은 장면의 렌더링에 있어서 렌더러(115)의 높은 정도의 유연성 및 적응성을 가능하게 할 수 있다. 이것은 예를 들어 렌더링된 오디오 장면을 적응시키고 맞춤화하기 위해 렌더러에 의해 사용될 수 있다. 예를 들어, 상이한 오디오 객체들의 상대 이득 및/또는 위치가 적응될 수 있고, 오디오 아이템들의 주파수 콘텐츠가 수정될 수 있고, 오디오 아이템들의 동적 범

위가 제어될 수 있고, 반향 특성들이 변경될 수 있고, 등등이다. 이에 따라, 렌더러(115)는 출력을 생성할 수 있으며, 여기서 오디오 장면은 사용되는 특정 재생 시스템에 대한 그리고/또는 청취자의 개인적 선호도에 대한 적응을 포함하여, 현재 응용/렌더링에 대한 특정 선호도에 적용된다. 이 접근법은 예를 들어 또한 오디오 장면 내의 가상 청취 위치에 있어서의 변경들에 대한 렌더링된 오디오 장면의 효율적인 로컬 적응을 허용할 수 있다. 예를 들어, 가상 현실 응용을 지원하기 위하여, 렌더러(115)는 사용자 위치 데이터를 동적으로 그리고 연속적으로 수신하고 오디오 장면 내의 사용자의 표시된 가상 위치에 있어서의 변경들에 응답하여 렌더링을 적응시킬 수 있다.

[0079] 렌더러(115)는 수신된 프레젠테이션 메타데이터에 기초하여 오디오 아이тем들을 렌더링하도록 배열된다. 특히, 프레젠테이션 메타데이터는 인코딩된/디코딩된 오디오 아이тем들의 렌더링의 가변 태양/특성/파라미터에 대한 제약들을 나타낼 수 있고, 렌더러(115)는 렌더링할 때 이러한 제약들을 따를 수 있다.

[0080] 렌더러(115)/렌더링 디바이스(109)로부터의 출력 오디오 신호들은 수신된 인코딩된 오디오 데이터 스트림으로부터 디코더(113)에 의해 생성된 디코딩된 오디오 아이тем들에 적용되는 렌더링 동작으로부터 기인한다. 렌더링 동작은, 외부에서 또는 로컬에서 적용될 수 있고 렌더링된 출력 오디오(의 태양들)에 지각적으로 영향을 미치는 몇몇 파라미터들을 가질 수 있다. 렌더링에 대한 프레젠테이션 제약들을 설명하는 프레젠테이션 메타데이터는 구체적으로 렌더링 파라미터들이 렌더링할 때 적용될 수 있는 세트(즉, 연속적으로 적응 가능한 파라미터들에 대해 값 범위, 또는 열거 파라미터들에 대해 이산 값들의 세트)를 제한하는 데이터일 수 있다.

[0081] 도 2는 오디오 인코더 유닛(103)의 요소들의 예를 더 상세히 도시한다. 이 예에서, 오디오 인코더 유닛(103)은 장면을 설명하는 입력 오디오 데이터를 수신하는 오디오 수신기(201)를 포함한다. 본 예에서, 오디오 장면은 3개의 상이한 유형의 오디오 데이터, 즉 채널 기반 오디오 아이тем들(C), 객체 기반 오디오 아이тем들(O), 및 장면 기반 오디오 아이тем들(S)에 의해 표현된다. 오디오 아이тем들은 임의의 적합한 형태일 수 있는 오디오 데이터에 의해 제공된다. 오디오 데이터는 예를 들어 오디오 아이тем들을 원시 WAV 파일들로서 또는 임의의 적합한 포맷에 따라 인코딩된 오디오로서 제공할 수 있다. 전형적으로, 입력 오디오 아이тем들은 높은 오디오 품질 및 높은 데이터 레이트에 있을 것이다.

[0082] 오디오 인코더 유닛(103)은 입력 오디오 아이тем들에 대한 프레젠테이션 메타데이터를 수신하도록 배열된 메타데이터 수신기(203)를 추가로 포함한다. 전송된 바와 같이, 프레젠테이션 메타데이터는 오디오 아이тем들의 렌더링에 대한 제약들을 제공할 수 있다.

[0083] 오디오 수신기(201) 및 메타데이터 수신기(203)는 수신된 오디오 아이тем들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하도록 배열된 오디오 인코더(205)에 결합된다. 오디오 인코더(205)는 이 예에서 구체적으로 인코딩된 오디오 아이тем들, 즉 인코딩된 오디오 데이터에 의해 표현되는 오디오 아이тем들을 생성한다. 입력 오디오 아이тем들에 대해, 출력/인코딩된 오디오 아이тем들은 또한 상이한 유형들의 오디오 아이тем들일 수 있고, 특정 예에서 구체적으로 채널 기반 오디오(C'), 객체 기반 오디오 아이тем들(O'), 및 장면 기반 오디오 아이тем들(S')일 수 있다.

[0084] 인코딩된 오디오 아이тем들 중 하나, 몇몇 또는 전부는 입력 오디오 아이тем들을 독립적으로 인코딩함으로써 생성될 수 있는데, 즉 인코딩된 오디오 아이тем은 인코딩된 입력 오디오 아이тем일 수 있다. 그러나, 몇몇 시나리오들에서, 인코딩된 오디오 아이тем들 중 하나 이상은 복수의 입력 오디오 아이тем을 표현하기 위해 생성될 수 있거나, 입력 오디오 아이тем은 복수의 인코딩된 오디오 아이тем들에서/에 의해 표현될 수 있다.

[0085] 많은 인코딩 알고리즘들 및 기술들이 알려져 있고 임의의 적합한 알고리즘, 표준, 및 접근법이 사용될 수 있다는 것이 인식될 것이다. 상이한 오디오 아이тем들에 대해 상이한 알고리즘들 및 기술들이 사용될 수 있다는 것이 또한 인식될 것이다. 예를 들어, 음악에 대응하는 오디오 아이тем은 AAC 인코딩 접근법을 사용하여 인코딩될 수 있고, 스피치에 대응하는 오디오 아이тем은 CELP 인코딩 접근법을 사용하여 인코딩될 수 있고, 등등이다. 인코딩된 포맷으로 이미 수신된 오디오 아이тем들에 대해, 오디오 인코더(205)에 의한 인코딩은 상이한 인코딩 포맷으로의 트랜스코딩일 수 있거나, 예를 들어 단순히 (예를 들어, 양자화 및/또는 클리핑 레벨들을 수정하는 것에 의한) 데이터 레이트 변환일 수 있다. 전형적으로, 인코딩은 비트 레이트 압축을 포함하고, 인코딩된 오디오 아이тем들은 입력 오디오 아이тем들보다 더 적은 비트들에 의해 표현된다.

[0086] 오디오 인코더 유닛(103)은 인코딩된 오디오 아이тем들에 대한 출력 프레젠테이션 메타데이터를 생성하도록 배열된 메타데이터 회로(207)를 추가로 포함한다. 프레젠테이션 메타데이터 회로(207)는 수신된 입력 프레젠테이션 메타데이터로부터 이러한 출력 프레젠테이션 메타데이터를 생성하도록 배열된다. 실제로, 많은 오디오 아이тем

들에 대해, 출력 프레젠테이션 메타데이터는 입력 프레젠테이션 메타데이터와 동일할 수 있다. 하나 이상의 오디오 아이টে에 대해, 출력 프레젠테이션 메타데이터는 나중에 더 상세히 설명될 바와 같이 수정될 수 있다.

- [0087] 오디오 인코더(205) 및 메타데이터 회로(207)는 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하도록 배열된 출력 회로(209)에 결합된다. 출력 회로(209)는 구체적으로 인코딩된 오디오 데이터 및 출력 메타데이터 둘 모두를 포함하는 인코딩된 오디오 데이터 스트림을 생성하는 비트스트림 패커(bitstream packer)일 수 있다. 인코딩된 오디오 데이터 스트림은 표준화된 포맷에 따라 생성될 수 있으며, 그에 의해 그것이 다양한 수신기들에 의해 해석될 수 있게 할 수 있다.
- [0088] 이에 따라, 출력 회로(209)는 비트레이트 감소된/인코딩된 오디오 아이টে들 및 출력 프레젠테이션 메타데이터를 받아들이는 비트스트림 패커로서 동작하고, 이들을 적합한 통신 채널을 통해, 이를테면 예를 들어 5G 네트워크를 통해 전달될 수 있는 비트스트림으로 결합한다.
- [0089] 도 3은 오디오 인코더 유닛(103)으로부터 인코딩된 오디오 데이터 스트림을 수신하고 처리할 수 있는 렌더링 디바이스(109)의 요소들의 특정 예를 예시한다. 렌더링 디바이스(109)는, 오디오 인코더 유닛(103)으로부터 인코딩된 오디오 데이터 스트림을 수신하고 수신된 데이터 스트림으로부터 상이한 데이터를 분리하는 비트스트림 언패커(bitstream unpacker)의 형태의 수신기(111)를 포함한다. 구체적으로, 수신기(111)는 인코딩된 오디오 아이টে들에 대한 개개의 오디오 데이터를 분리하고 이들을 디코더(113)에 공급할 수 있다.
- [0090] 디코더(113)는 구체적으로 수신된 인코딩된 오디오 아이টে들을 디코딩하여 전형적으로 채널, 객체, 및 장면 기반 오디오 아이টে들의 인코딩되지 않은 표현들을 생성하도록 배열된다.
- [0091] 많은 오디오 아이টে들에 대해, 디코더(113)는 오디오 인코더(205)에 의해 수행되는 인코딩을 반전시킬 수 있다. 다른 오디오 아이টে들에 대해, 디코딩은 예를 들어 인코딩 동작을 부분적으로만 반전시킬 수 있다. 예를 들어, 오디오 인코더(205)가 오디오 아이টে들을 단일의 결합된 오디오 아이টে으로 결합한 경우, 디코더(113)는 결합된 오디오 아이টে만을 디코딩하고 개개의 오디오 아이টে들을 완전히 생성하지는 않을 수 있다. 개개의 실시예의 특정 선호들 및 요건들에 따라 임의의 적합한 디코딩 알고리즘 및 기술이 사용될 수 있다는 것이 인식될 것이다.
- [0092] 디코딩된 오디오 아이টে들은 예를 들어 바이노럴 신호 또는 서라운드 사운드 신호로서 전송된 바와 같은 오디오 아이টে들을 렌더링함으로써 오디오 장면을 렌더링하도록 배열된 렌더러(115)에 공급된다.
- [0093] 렌더링 디바이스(109)는 수신기(111)로부터 프레젠테이션 메타데이터를 공급받는 메타데이터 컨트롤러/회로(301)를 추가로 포함한다. 이 예에서, 메타데이터 컨트롤러(301)는 또한 예를 들어 로컬 선호들 또는 요건들, 이를테면 예를 들어 사용되는 재생 시스템의 개개의 사용자 선호들 또는 특성들을 반영할 수 있는 로컬 프레젠테이션 메타데이터를 수신할 수 있다.
- [0094] 이에 따라, 수신된 비트스트림으로부터 언패킹된 오디오 프레젠테이션 메타데이터에 더하여, 렌더링 디바이스(109)는 또한 예를 들어 하나 이상의 입력 인터페이스를 통해 제공될 수 있는 로컬 오디오 프레젠테이션 메타데이터를 받아들일 수 있다. 이러한 데이터는 예를 들어 다음과 같은, 인코더 측에서 이용 가능하지 않은 오디오가 제시될 컨텍스트에 관한 정보를 제공할 수 있다:
- [0095] - 원하는 프레젠테이션 (라우트스피커) 구성;
 - [0096] - 사용자 선호(예를 들어, 가상 회의에서의 참가자들의 오디오의 오디오 레벨들 및 배향들);
 - [0097] - 예를 들어 룸의 반향과 같은, 로컬 음향 시설의 특성들. 이것은 렌더러가 어느 분위기 효과들 및 특성들을 오디오 아이টে들에 적용할지를 결정할 수 있게 할 수 있다;
 - [0098] - 로컬 오디오 신호들(예를 들어, 오디오 아이টে들에 대한 이득을 선택할 때 고려하기 위해);
 - [0099] - 청취자 위치; 및
 - [0100] - 청취자 머리 배향.
- [0101] 메타데이터 컨트롤러(301)는 수신된 메타데이터와 로컬 메타데이터를 함께 병합하고 그것을 렌더러(115)에 제공할 수 있으며, 이 렌더러는 프레젠테이션 메타데이터의 제약들에 따라 오디오 아이টে들을 렌더링하도록 진행할 수 있다.
- [0102] 렌더러(115)는 디코더(113)에 의해 생성된 오디오 아이টে들(C", O", 및 S")을 원하는 프레젠테이션 구성(예를

들어, 바이노럴 또는 서라운드 사운드)의 제시 가능한 오디오로 결합할 수 있다.

- [0103] 렌더러(115)는 구체적으로 메타데이터 컨트롤러(301)로부터 수신된 메타데이터에 따라, 그리고 수신된 프레젠테이션 메타데이터의 제약들에 의해 제약되는, 즉 인코더 측으로부터 제약되는 렌더링된 오디오에 따라 오디오 프레젠테이션을 생성할 수 있다. 이것은 클라이언트 측에서 얼마간의 유연성을 여전히 허용하면서 오디오 렌더링 및 제시된 오디오 장면에 대한 소스 측/콘텐츠 제공자 제어를 제공한다. 이것은 예를 들어 콘텐츠 창조자가 소정의 제한된 제어를 최종 사용자 등에게 제공하도록 설계된 몰입형 응용의 제어를 유지하는 서비스 또는 응용을 제공하는 데 사용될 수 있다.
- [0104] 보다 상세하게, 메타데이터 컨트롤러(301)는 수신된 메타데이터, 예를 들어 오디오 아이템들의 억제, 따라서 로컬 메타데이터를 처리할 수 있다. 메타데이터 컨트롤러(301)는 예를 들어 로컬 메타데이터, 예를 들어 회전 또는 고도의 범위, 따라서 수신된 메타데이터를 제한할 수 있다.
- [0105] 몇몇 실시예들에서, 렌더러(115)는 렌더링 디바이스(109)와는 상이한 디바이스 또는 기능 엔티티일 수 있다. 예를 들어, 예상된 3GPP IVAS 코덱과 같은 표준이 디코더(113)의 동작을 규정할 수 있지만, 렌더러(115)가 독점적이고 더 자유롭게 적응 가능하도록 허용할 수 있다. 몇몇 실시예들에서, 메타데이터 컨트롤러(301)는 상이한 디바이스 또는 기능 엔티티의 일부일 수 있다.
- [0106] 그러한 실시예에서, 외부 렌더러는 이에 따라 디코딩된 오디오 아이템들(0", C", S") 및 수신된 프레젠테이션 메타데이터를 처리 및 해석하도록 요구된다. 외부 렌더러에 의한 렌더링 동작은 프레젠테이션 메타데이터에 의해 제공되는 제약들에 여전히 따라야 한다.
- [0107] 이에 따라 프레젠테이션 메타데이터는 클라이언트들에서의 렌더링 동작을 제어하기 위해 소스 측/콘텐츠 제공자에 의해 사용되는 데이터일 수 있다. 렌더링은 프레젠테이션 메타데이터에 따라 적응/제한되어야 한다.
- [0108] 그러나, 클라이언트 측 렌더러(115)에 의한 렌더링을 제어하는 데 사용되는 프레젠테이션 메타데이터에 더하여, 오디오 인코더 유닛(103)의 오디오 인코더(205)는 또한 입력 프레젠테이션 메타데이터에 응답하여 인코딩을 적응시키도록 배열된다. 입력 프레젠테이션 메타데이터는 오디오 인코더(205)에 공급되고 이것은 하나 이상의 오디오 아이템의 인코딩을, (전형적으로 그러한 하나 이상의 오디오 아이템에 대한) 프레젠테이션 메타데이터에 기초하여 수정할 수 있다. 이에 따라 오디오 인코더(205)는 오디오 아이템들과 함께 수신된 프레젠테이션 메타데이터에 응답하는 적응 가능한 인코더이다.
- [0109] 오디오 인코더(205)는 구체적으로 오디오 아이템들의 인코딩을 수행하도록 배열된 인코딩 회로(211), 및 프레젠테이션 메타데이터에 기초하여 인코딩 회로(211)에 의한 인코딩을 적응시키도록 배열된 인코딩 어댑터(213)를 포함한다.
- [0110] 인코딩 어댑터(213)는 주어진 오디오 아이템에 대한 인코딩의 파라미터를, 그러한 오디오 아이템에 대한 프레젠테이션 메타데이터에 기초하여 설정하도록 배열될 수 있다. 예를 들어, 그것은 인코딩을 위한 비트 레이트 할당/타겟, 양자화 레벨, 마스킹 임계치, 주파수 범위 등을, 예를 들어 오디오 아이템에 대해 허용 가능하도록 프레젠테이션 메타데이터에 의해 표시된 이득 범위 또는 위치 범위에 기초하여 설정하도록 배열될 수 있다.
- [0111] 많은 실시예들에서, 인코딩 회로(211)는 수신된 입력 오디오 아이템들에 비해 감소된 수의 비트들로 오디오 아이템들을 인코딩하도록 배열된 비트레이트 압축기이다. 따라서 인코딩은 비트레이트 압축될 수 있고, 그에 의해 더 효율적이고 분배하기가 더 쉬운 인코딩된 오디오 데이터 스트림이 생성될 수 있게 할 수 있다. 그러한 실시예들에서, 인코딩 어댑터(213)는 (적합한 최적화 기준/알고리즘에 따라 렌더링된 오디오의 품질을 최적화하기 위해) 프레젠테이션 메타데이터에 기초하여 인코딩 회로(211)의 비트레이트 감소를 적응시킬 수 있다.
- [0112] 인코딩 어댑터(213)는 예를 들어, 프레젠테이션 메타데이터를 분석하고 다양한 입력 오디오 아이템들의 비트레이트 감소를 최상으로 수행하는 방법에 대한 결정들을 행하는 코딩 분석 프로세스를 실행할 수 있다. 인코딩 어댑터(213)에 의해 수행될 수 있는 동작들 및 적응의 예들은 다음을 포함한다:
- [0113] - 비트레이트 감소를 위해 고수할 인코딩 회로(211)에 대한 (최소) 마스킹 레벨들을 시그널링함. 인코딩 어댑터(213)는 어느 오디오 아이템들이 그리고 어느 레벨들에서 그리고 어느 방향에서 공동-제시되는지에 관한 정보를 갖는다. 이것은 그것이 그후에 인코딩에 의해 사용되는 마스킹 레벨들로 개개의 오디오 아이템들에 대한 마스킹 레벨을 적응시킬 수 있게 할 수 있다.
- [0114] - 오디오 아이템들을 변환함, 예를 들어 오디오 객체들을 채널 또는 장면 기반 오디오로 바꿈.

- [0115] - (연관된 업믹스 파라미터들을 갖는) 다운믹싱을 위한 오디오 아이템들을 선택함, 여기서 다운믹스는 파라미터 다운믹스 코딩의 아티팩트들이 함께 제시되는 다양한 오디오 아이템들에 의해 충분히 마스킹되는 것을 보장하면서 디코더 측에서 몰입형 오디오를 재구성하기 위해 업믹싱될 수 있음. 추가의 개선들로서, 인코딩 어댑터 (213)는 다음을 할 수 있다
- [0116] - 최대 성능/최소 아티팩트들을 위해 다운믹싱/업믹싱 이득들을 최적화함;
- [0117] - 최적의 시간/주파수 특성들을 갖는 업믹싱 파라미터들을 선택함.
- [0118] - 오디오 아이템들을 결합된 오디오 아이템들로 비가역적으로 결합함, 결합된 오디오 아이템들은 이어서 렌더러 (115)에 의해 단일 오디오 아이템으로서 렌더링될 수 있음. 이것은 모든 오디오 정보를 렌더링 측에서 개별적으로 이용 가능하게 할 고유의 필요성이 없다는 것을 이용할 수 있다. 예를 들어, 몇몇 입력 오디오 아이템들의 별개의 적응이 허용되지 않는 경우(예를 들어, 그들이 동일한 위치로 렌더링되도록 요구될 수 있음), 오디오 아이템들을 개별적으로 이용 가능하게 하는 것이 필요하지 않다. 예를 들어, 유사한 배향 및 이득 적응 제약들을 갖는 다수의 입력 오디오 객체들이 하나의 장면 기반 오디오 아이템으로 결합될 수 있으며, 여기서 렌더링 동안 장면에 대한 이득 및 배향은 전체로서 여전히 적용될 수 있지만, 전자의 객체들은 장면에서 고정된 상대 오디오 레벨들 및 고정된 상대 위치들을 가질 것이다.
- [0119] - 오디오 아이템들에 대한 프레젠테이션 메타데이터에 의존하여 상이한 오디오 아이템들에 상이한 비트레이트 예산들을 할당함. 예를 들어, 비트레이트는 그들이 각각 나타내는 마스킹되지 않은 정보의 양에 기초하여 오디오 아이템들에 할당될 수 있다.
- [0120] 이어서 인코딩 회로(211)는 인코딩 어댑터(213)에 의해 생성된 코딩 제어 데이터에 따라 오디오 아이템들의 코딩을 채용할 수 있다. 예를 들어, 인코딩 회로(211)는 몇몇 채널, 객체 및 장면 기반 오디오 아이템들의 비트레이트 감소된(예를 들어, 양자화된, 파라미터화된 ...) 버전들을 생성할 수 있다. 게다가, 예를 들어 상이한 오디오 아이템들의 인코딩의 일부로서의 결합 또는 변환으로 인해, 인코딩된 오디오 아이템들 중 적어도 일부가 입력 오디오 아이템들과는 상이한 오디오 정보를 표현할 수 있는데, 즉 입력 오디오 아이템들과 인코딩된 오디오 아이템들 사이에 직접적인 대응이 존재하지 않을 수 있다.
- [0121] 몇몇 실시예들에서, 오디오 인코더(205)는 구체적으로 입력 오디오 아이템들을 하나 이상의 결합된 오디오 아이템들로 결합하도록 배열된 결합기(215)를 포함할 수 있다. 결합기(215)는 구체적으로 제1 및 제2 입력 오디오 아이템을 결합된 오디오 아이템으로 결합할 수 있다. 이어서 결합된 오디오 아이템은 인코딩되어 결합된 인코딩된 오디오 아이템을 생성할 수 있고, 이러한 결합된 인코딩된 오디오 아이템은 인코딩된 오디오 데이터 스트림에 포함되어, 전형적으로 제1 및 제2 오디오 아이템을 대체할 수 있다. 이에 따라, 제1 및 제2 오디오 아이템들을 개별적으로 인코딩하기보다는, 결합기(215)는 그들을 단일의 인코딩된 오디오 아이템으로 결합할 수 있고, 이는 이어서 인코딩된 오디오 데이터 스트림에 포함되는 반면, 제1 또는 제2 오디오 아이템들 각각에 대해 개개의 인코딩된 오디오 데이터가 포함되지 않는다.
- [0122] 오디오 아이템들의 결합은 수신된 프레젠테이션 메타데이터에 응답하여 수행된다. 많은 실시예들에서, 결합을 위해 선택되는 오디오 아이템들은 프레젠테이션 메타데이터에 기초하여 선택된다. 예를 들어, 인코딩 어댑터 (213)는 오디오 아이템들에 대한 제약들이 유사성 기준을 충족시키는 요건을 포함하는 기준에 응답하여 결합을 위해 오디오 아이템들을 선택할 수 있다.
- [0123] 예를 들어, 결합될 오디오 아이템들에 대해, 프레젠테이션 메타데이터에 의해 표시된 바와 같은 오디오 아이템들에 대한 제약들이 모순되지 않아야 하는 것이, 즉 둘 모두의 제약들을 충족시키는 것이 가능해야 하는 것이 요구될 수 있다. 이에 따라, 프레젠테이션 메타데이터에 의해 표시된 제약들이 모순되지 않고, 예를 들어 제약들은 결합되는 둘 모두의(또는 모든) 오디오 아이템들에 대한 렌더링 제약들이 충족될 수 있게 하는 적어도 하나의 렌더링 파라미터 값이 존재하도록 적어도 중첩을 갖는 것이 요구될 수 있다. 인코딩 어댑터(213)는 프레젠테이션 메타데이터가 공통 렌더링 파라미터에 대한 양립 불가능한 제약들을 설명하지 않을 것을 요구할 수 있다.
- [0124] 예를 들어, 프레젠테이션 메타데이터는 오디오 장면 내의 오디오 아이템들의 위치에 대한 제약들을 설명할 수 있다. 그러한 경우에, 위치 제약들은 중첩되어야 하고 몇몇 공통 허용 가능 위치들이 존재해야 하는 것이 요구될 수 있다.
- [0125] 결합할 오디오 아이템들의 선택은 오디오 아이템들에 대한 프레젠테이션 메타데이터에 기초할 수 있다. 이에

따라, 결합을 위한 제1 및 제2 오디오 아이템의 선택은 제1 및 제2 오디오 아이템들에 대한 프레젠테이션 메타데이터에 기초할 수 있다. 예를 들어, 위에서 언급된 바와 같이, 제1 및 제2 오디오 아이템들에 대한 프레젠테이션 메타데이터가 모순되는 제약들을 정의하지 않을 것이 요구될 수 있다.

- [0126] 몇몇 실시예들에서, 제1 및 제2 오디오 아이템들은 예를 들어 예컨대 가장 유사한 동일 파라미터에 대한 제약들을 갖는 오디오 아이템들이도록 선택될 수 있다. 예를 들어, 실질적으로 동일한 위치 제약들을 갖는 오디오 아이템들이 선택될 수 있다.
- [0127] 구체적으로, 2개의 오디오 아이템에 대한 유사성 척도가 허용 가능한 위치들 사이의 중첩을 반영하도록 결정될 수 있다. 예를 들어, 유사성 척도는 2개의 오디오 아이템에 대한 개개의 허용 가능 위치들의 볼륨들의 합에 대한 중첩하는 허용 가능 위치들의 영역의 볼륨 사이의 비율로서 생성될 수 있다.
- [0128] 다른 예로서, 그들의 위치 적응 제약들에 대한 유사성 기준을 충족시키는 다수의 오디오 객체들은, 각자의 위치 범위들 또는 공간 볼륨들이 중첩하지 않을 수 있을지라도, 장면 기반 오디오 아이템으로 결합될 수 있으며, 여기서 오디오 소스들은 그때부터 장면 기반 오디오에서 서로에 대해 고정된 상대 배향들을 가질 것이지만(즉, 개별적으로 적응 가능하지 않음), 그들의 배향들은 여전히 전체로서 함께 적응될 수 있다.
- [0129] 다른 예로서, 유사성 척도는 2개의 오디오 아이템에 대한 중첩 이득 범위의 크기를 반영하도록 생성될 수 있다. 공통 허용 가능 이득 범위가 클수록, 유사성이 크다.
- [0130] 인코딩 어댑터(213)는 오디오 아이템들의 상이한 쌍들에 대해 그러한 유사성 척도들을 평가하고 예를 들어 유사성 척도들이 주어진 임계치보다 더 높은 쌍들을 선택할 수 있다. 이어서 이러한 오디오 아이템들은 단일의 결합된 오디오 아이템들로 결합될 수 있다.
- [0131] 많은 실시예들에서, 인코딩 어댑터(213)는 입력 프레젠테이션 메타데이터로부터 결합된 오디오 아이템에 대한 결합된 프레젠테이션 메타데이터를 생성하도록 추가로 배열된다. 이어서 이러한 프레젠테이션 메타데이터는 비트스트림 패커(209)에 공급되고, 이 비트스트림 패커는 그것을 출력 인코딩된 오디오 데이터 스트림에 포함시킨다.
- [0132] 메타데이터 회로(207)는 구체적으로 결합된 오디오 아이템에 링크되고 결합된 프레젠테이션 메타데이터에 대한 렌더링 제약들을 제공하는 결합된 프레젠테이션 메타데이터를 생성할 수 있다. 이어서 연관된 결합된 프레젠테이션 메타데이터를 갖는 생성된 결합된 오디오 아이템은 임의의 다른 오디오 아이템으로서 취급될 수 있고, 실제로 클라이언트/디코더/렌더러는 결합된 오디오 아이템이 실제로 오디오 인코더(205)에 의한 입력 오디오 아이템들의 결합에 의해 생성되는 것을 인식하지 못할 수도 있다. 오히려, 결합된 오디오 아이템 및 연관된 프레젠테이션 메타데이터는 클라이언트 측에 대해 입력 오디오 아이템들 및 연관된 프레젠테이션 메타데이터와 구별 불가능할 수 있고, 임의의 다른 오디오 아이템으로서 렌더링될 수 있다.
- [0133] 많은 실시예들에서, 결합된 프레젠테이션 메타데이터는 예를 들어 결합된 오디오 아이템에 대한 프레젠테이션 파라미터에 대한 제약을 반영하도록 생성될 수 있다. 제약은 그것이 이러한 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 바와 같이 결합되는 오디오 아이템들에 대한 개개의 제약들을 충족시키도록 결정될 수 있다. 구체적으로, 제1 및 제2 오디오 아이템에 대한 결합된 오디오 아이템에 대한 제약은 제1 오디오에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제1 오디오 아이템에 대한 제약 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제2 오디오 아이템에 대한 제약 둘 모두를 충족시키는 제약으로서 결정될 수 있다. 이에 따라, 결합된 프레젠테이션 메타데이터는 결합된 제약이 충족되면 개별 오디오 아이템들에 대한 개별 제약들이 충족되는 것을 보장하는 하나 이상의 제약을 제공하도록 생성된다.
- [0134] 예를 들어, 오디오 객체인 제1 오디오 아이템에 대해, 입력 프레젠테이션 메타데이터는 그것이, 말하자면, -6dB 내지 0dB의 범위 내의 상대 이득으로, 그리고, 말하자면, ([0,100], [-40,60], [0.5,1.5])의 (방위각, 고도, 반경)의 좌표 볼륨 내의 위치에 렌더링되어야 함을 나타낼 수 있다. 오디오 객체인 제2 오디오 아이템에 대해, 입력 프레젠테이션 메타데이터는 그것이, 말하자면, -3dB 내지 3dB의 범위 내의 상대 이득으로, 그리고, 말하자면, ([-100,80], [-20,70], [0.2,1.0])의 (방위각, 고도, 반경)의 좌표 볼륨 내의 위치에 렌더링되어야 함을 나타낼 수 있다. 이 경우에, 결합된 프레젠테이션 메타데이터는 오디오 객체인 결합된 오디오 아이템이, 말하자면, -3dB 내지 0dB의 범위 내의 상대 이득으로, 그리고, 말하자면, ([0,80], [-20,60], [-0.5,1.0])의 (방위각, 고도, 반경)의 좌표 볼륨 내의 위치에 렌더링되어야 함을 나타내도록 생성될 수 있다. 이것은 결합된 오디오 아이템이 제1 및 제2 오디오 아이템 둘 모두에 대해 허용 가능할 방식으로 렌더링되는 것을 보장할 것이다.
- [0135] 몇몇 실시예들에서, 오디오 인코더(205)는 다른 오디오 아이템에 대한 프레젠테이션 메타데이터에 기초하여 하

나의 오디오 아이템의 압축을 적응시키도록 배열될 수 있다.

- [0136] 낮은 복잡도 예로서, 하나의 오디오 아이템의 압축은 다른 오디오 아이템에 대한 근접도 및 이득/레벨에 의존할 수 있다. 예를 들어, 현재 오디오 아이템에 대한 프레젠테이션 메타데이터가 위치 범위 및 레벨 범위를 표시하는 경우, 이것은 제2 오디오 아이템에 대한 위치 범위 및 레벨 범위와 비교될 수 있다. 제2 오디오 아이템이 제1 오디오 아이템에 가깝게 위치되도록 제약되고 제1 오디오 아이템보다 상당히 더 높은 레벨에서 렌더링되도록 제약되는 경우, 제1 오디오 아이템은 청취자에 의해 단지 미미하게 인지될 가능성이 있다. 따라서, 제1 오디오 아이템의 인코딩은 다른 오디오 아이템이 존재하지 않는 경우보다 더 높은 압축/비트 레이트 감소를 가질 수 있다. 구체적으로, 제1 오디오 아이템의 인코딩을 위한 비트 레이트 할당은 하나 이상의 다른 오디오 아이템까지의 거리 및 그의 레벨에 의존할 수 있다.
- [0137] 몇몇 실시예들에서, 인코딩 어댑터(213)는 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록 배열될 수 있다. 마스킹 효과는 제2 오디오 아이템의 렌더링으로부터 제1 오디오 아이템에 도입된 마스킹의 정도를 나타내는 마스킹 척도에 의해 표현될 수 있다. 이에 따라 마스킹 척도는 제2 오디오 아이템의 존재 시의 제1 오디오 아이템의 지각적 중요성을 나타낸다.
- [0138] 마스킹 척도는 구체적으로 제2 오디오 아이템이 프레젠테이션 메타데이터에 의해 표시된 제약들에 따라 렌더링될 때 제1 오디오 아이템으로부터의 수신된 사운드의 레벨에 대한 제2 오디오 아이템으로부터의 수신된 사운드의 레벨의 표시로서 생성될 수 있다.
- [0139] 예를 들어, 그의 최고 이득에서의 제2 오디오 아이템에 대한 그의 최저 이득에서의 제1 오디오 아이템의 마스킹 효과가 제2 오디오 아이템의 마스킹 레벨을 추정하기 위해 취해질 수 있고, 그 반대도 마찬가지이다.
- [0140] 다른 예로서, 제1 오디오 아이템과 제2 오디오 아이템 사이의 가장 먼(또는 예를 들어, 평균) 거리가 결정될 수 있고 그들 사이의 감쇠가 추정될 수 있다. 이어서 감쇠에 대한 보상 후의 상대 레벨 차이에 기초하여 마스킹 효과가 추정될 수 있다.
- [0141] 다른 예로서, 시스템이 공칭 청취 위치를 채용하는 경우, 제1 및 제2 오디오 아이템들 각각으로부터의 청취 위치에서의 신호 레벨은 상대 이득 레벨들 또는 신호 레벨들 및 사운드 소스들의 위치들로부터의 감쇠에 있어서의 차이에 기초하여 결정될 수 있다. 오디오 아이템 위치들은, 예를 들어, 마스킹 효과가 최소화되도록 허용 가능 위치들로부터 선택될 수 있다(제1 오디오 아이템에 대한 가장 가까운 허용 가능 위치 및 제2 오디오 아이템에 대한 가장 먼 위치).
- [0142] 이에 따라, 인코딩 어댑터(213)는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시되는 제2 오디오 아이템에 대한 이득/레벨 제약 및 위치 제약에 기초하여; 그리고 종종 또한 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시되는 제1 오디오 아이템에 대한 이득/레벨 제약 및 위치 제약에 기초하여 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정할 수 있다.
- [0143] 몇몇 실시예들에서, 인코딩 어댑터(213)는 제2 오디오 아이템에 대한 프레젠테이션 메타데이터에 기초하여 제1 오디오 아이템에 대한 마스킹 임계치를 직접 결정할 수 있고, 인코딩 회로(211)는 결정된 마스킹 임계치를 사용하여 제1 오디오 아이템을 인코딩하도록 진행할 수 있다.
- [0144] 몇몇 실시예들에서, 오디오 인코더(205)에 의한 인코딩의 적응은 어떠한 다른 기능도 그에 맞춰 적응되지 않는 내부 프로세스일 수 있다. 예를 들어, 복수의 오디오 아이템들의 결합된 오디오 아이템으로의 비가역적 결합이, 결합된 오디오 아이템이 인코딩된 오디오 데이터 스트림에 포함되고 결합된 오디오 아이템이 어떻게 생성되었는지에 대한 표시 없이, 즉 렌더링 디바이스가 결합된 오디오 아이템의 임의의 특정 처리를 수행함이 없이 수행될 수 있다.
- [0145] 그러나, 많은 실시예들에서, 오디오 인코더(205)는 입력 프레젠테이션 메타데이터에 응답하여 인코딩이 어떻게 적응되는지를 나타내는 인코딩 적응 데이터를 생성할 수 있다. 이어서 이러한 인코딩 적응 데이터는 인코딩된 오디오 데이터 스트림에 포함될 수 있다. 이러한 접근법에서, 렌더링 디바이스(109)는 그에 따라서 인코딩 적응의 정보를 가질 수 있고 디코딩 및/또는 렌더링을 그에 맞춰 적응시키도록 배열될 수 있다.
- [0146] 예를 들어, 오디오 인코더(205)는 음향 환경 데이터의 어느 오디오 아이템들이 실제로 결합된 오디오 아이템들 인지를 표시하는 데이터를 생성할 수 있다. 그것은 결합의 몇몇 파라미터들을 추가로 표시할 수 있고, 실제로 이들은 많은 실시예들에서 렌더링 디바이스(109)가 결합된 오리지널 오디오 아이템들의 표현들을 생성할 수 있게 할 수 있다. 실제로, 몇몇 실시예들에서, 결합된 오디오 아이템은 입력 오디오 아이템들의 다운믹스로서 생

성될 수 있고 오디오 인코더(205)는 파라미터 업믹스 데이터를 생성하고 이것을 인코딩된 오디오 데이터 스트림에 포함시킬 수 있으며, 그에 의해 렌더링 디바이스가 합리적인 업믹싱을 수행하는 것을 가능하게 할 수 있다.

[0147] 다른 예로서, 보통 말하는 그런 디코딩이 적용되지 않을 수 있지만, 정보는 청취자/최종 사용자와의 상호작용에서 사용될 수 있다. 예를 들어, 그들의 적응 제약들에서 '가까운' 것으로 간주되는 다수의 오디오 객체들은 인코더에 의해 하나의 장면 기반 오디오 아이템으로 결합될 수 있는 반면, '가상 객체들'로서의 그들의 존재는 인코딩 적응 데이터에서 디코더에 시그널링된다. 이어서 사용자는 이러한 정보를 제공받을 수 있고, 가상 객체들에 대한 캐리어로서 장면 기반 오디오를 통지받기/인식하기보다는 '가상 사운드 소스들'(그들이 장면 기반 오디오에서 결합되었기 때문에 단지 전체로서이긴 하지만)을 수동으로 제어하도록 제안받을 수 있다.

[0148] 몇몇 실시예들에서, 프레젠테이션 메타데이터는 하나 이상의 오디오 아이템에 대한 우선순위 데이터를 포함할 수 있고, 오디오 인코더(205)는 제1 오디오 아이템에 대한 우선순위 표시에 응답하여 제1 오디오 아이템에 대한 압축을 적응시키도록 배열될 수 있다.

[0149] 우선순위 표시는 오디오 장면 내의 오디오 아이템의 지각적 중요성 또는 중대성을 나타내는 렌더링 우선순위 표시일 수 있다. 예를 들어, 그것은 메인 스피커를 나타내는 오디오 아이템이, 말하자면, 배경 내의 새 지저귀음을 나타내는 오디오 아이템보다 더 중요하다는 것을 나타내는 데 사용될 수 있다.

[0150] 렌더러(115)는 우선순위 표시들에 기초하여 렌더링을 적응시킬 수 있다. 예를 들어, 청력이 감소된 청취자들에게 대해, 렌더러(115)는 낮은 우선순위 배경 잡음에 비해 높은 우선순위 메인 대화에 대한 이득을 증가시킬 수 있고, 그에 의해 스피치를 더 이해하기 쉽게 만들 수 있다.

[0151] 게다가, 오디오 인코더(205)는 우선순위를 감소시키기 위해 압축을 증가시킬 수 있다. 예를 들어, 오디오 아이템들을 결합하기 위해, 우선순위 레벨들이 주어진 레벨 미만이어야 하는 것이 요구될 수 있다. 다른 예로서, 오디오 인코더(205)는 우선순위 레벨이 주어진 레벨 미만인 모든 오디오 아이템들을 결합할 수 있다.

[0152] 몇몇 실시예들에서, 각각의 오디오 아이템들에 대한 비트 할당은 우선순위 레벨에 의존할 수 있다. 예를 들어, 상이한 오디오 아이템들에 대한 비트 할당은 우선순위를 포함한 복수의 파라미터를 고려하는 알고리즘 또는 공식에 기초할 수 있다. 주어진 오디오 아이템에 대한 비트 할당은 우선순위가 증가함에 따라 단조 증가할 수 있다.

[0153] 위의 설명은 명료함을 위해 상이한 기능 회로들, 유닛들 및 프로세서들을 참조하여 본 발명의 실시예들을 설명하였음을 알 것이다. 그러나, 본 발명으로부터 벗어남이 없이 상이한 기능 회로들, 유닛들 또는 프로세서들 간의 기능의 임의의 적합한 분산이 사용될 수 있음이 명백할 것이다. 예를 들어, 별개의 프로세서들 또는 컨트롤러들에 의해 수행되도록 예시된 기능이 동일한 프로세서 또는 컨트롤러에 의해 수행될 수 있다. 따라서, 특정 기능 유닛들 또는 회로들에 대한 참조들은 오로지 엄격한 논리적 또는 물리적 구조 또는 조직을 나타내기보다는 설명된 기능을 제공하기 위한 적합한 수단에 대한 참조들로 간주되어야 한다.

[0154] 본 발명은 하드웨어, 소프트웨어, 펌웨어 또는 이들의 임의의 조합을 포함한 임의의 적합한 형태로 구현될 수 있다. 본 발명은 선택적으로 하나 이상의 데이터 프로세서 및/또는 디지털 신호 프로세서 상에서 실행되는 컴퓨터 소프트웨어로서 적어도 부분적으로 구현될 수 있다. 본 발명의 실시예의 요소들 및 컴포넌트들은 임의의 적합한 방식으로 물리적으로, 기능적으로 그리고 논리적으로 구현될 수 있다. 실제로, 기능은 단일 유닛에서, 복수의 유닛에서 또는 다른 기능 유닛들의 일부로서 구현될 수 있다. 그렇기 때문에, 본 발명은 단일 유닛에서 구현될 수 있거나, 상이한 유닛들, 회로들 및 프로세서들 간에 물리적으로 그리고 기능적으로 분산될 수 있다.

[0155] 일반적으로, 오디오 인코딩 장치, 오디오를 인코딩하는 방법, 및 방법을 구현하는 컴퓨터 프로그램 제품의 예들이 아래의 실시예들에 의해 명시된다.

[0156] 1. 오디오 인코딩 장치로서,

[0157] 오디오 장면을 나타내는 복수의 오디오 아이템들을 수신하기 위한 오디오 수신기(201);

[0158] 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하기 위한 메타데이터 수신기(203)로서, 입력 프레젠테이션 메타데이터는 복수의 오디오 아이템들의 렌더링에 대한 프레젠테이션 제약들을 설명하는, 상기 메타데이터 수신기(203);

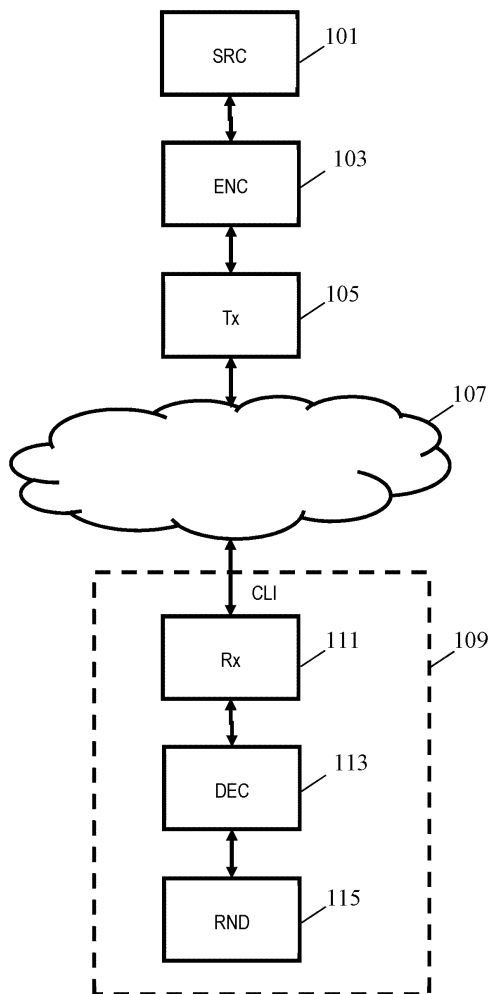
[0159] 복수의 오디오 아이템들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하기 위한 오디오 인코더(205)로서, 인코딩은 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 오디오 인코더(205);

- [0160] 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하기 위한 메타데이터 회로(207); 및
- [0161] 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하기 위한 출력 회로(209)를 포함하는, 오디오 인코딩 장치.
- [0162] 2. 제1항에 있어서, 오디오 인코더(205)는 복수의 오디오 아이템들 중 적어도 제1 오디오 아이템과 제2 오디오 아이템을 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 결합함으로써 결합된 오디오 아이템을 생성하기 위한 결합기(215)를 포함하고, 오디오 인코더(205)는 결합된 오디오 아이템을 인코딩함으로써 제1 및 제2 오디오 아이템에 대한 결합된 오디오 인코딩 데이터를 생성하도록, 그리고 결합된 오디오 인코딩 데이터를 인코딩된 오디오 데이터에 포함시키도록 배열되는, 오디오 인코딩 장치.
- [0163] 3. 제2항에 있어서, 결합기(215)는 제1 오디오 아이템 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 복수의 오디오 아이템들로부터 제1 오디오 아이템 및 제2 오디오 아이템을 선택하도록 배열되는, 오디오 인코딩 장치.
- [0164] 4. 제2항 또는 제3항에 있어서, 결합기(215)는 적어도 몇몇의 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터와 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터가 유사성 기준을 충족시킨다는 결정에 응답하여 제1 오디오 아이템 및 제2 오디오 아이템을 선택하도록 배열되는, 오디오 인코딩 장치.
- [0165] 5. 제2항 내지 제4항 중 어느 한 항에 있어서, 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터는 이득 제약 및 위치 제약 중 적어도 하나를 포함하는, 오디오 인코딩 장치.
- [0166] 6. 제2항 내지 제5항 중 어느 한 항에 있어서, 오디오 인코더(205)는 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 결합된 오디오 아이템에 대한 결합된 프레젠테이션 메타데이터를 생성하도록, 그리고 결합된 프레젠테이션 메타데이터를 출력 프레젠테이션 메타데이터에 포함시키도록 추가로 배열되는, 오디오 인코딩 장치.
- [0167] 7. 제6항에 있어서, 오디오 인코더(205)는 결합된 오디오 아이템에 대한 프레젠테이션 파라미터에 대한 제약을 반영하기 위해 적어도 몇몇의 결합된 프레젠테이션 메타데이터를 생성하도록 배열되고, 제약은 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제1 오디오 아이템에 대한 제약 및 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제2 오디오 아이템에 대한 제약 둘 모두를 충족시키는 제약으로서 결정되는, 오디오 인코딩 장치.
- [0168] 8. 제1항 내지 제7항 중 어느 한 항에 있어서, 오디오 인코더(205)는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.
- [0169] 9. 제8항에 있어서, 오디오 인코더(205)는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록, 그리고 마스킹 효과에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.
- [0170] 10. 제9항에 있어서, 오디오 인코더(205)는 제2 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 의해 표시된 제2 오디오 아이템에 대한 이득 제약 및 위치 제약 중 적어도 하나에 응답하여 제2 오디오 아이템으로부터의 제1 오디오 아이템에 대한 마스킹 효과를 추정하도록 배열되는, 오디오 인코딩 장치.
- [0171] 11. 제8항 내지 제10항 중 어느 한 항에 있어서, 오디오 인코더(205)는 제1 오디오 아이템에 대한 입력 프레젠테이션 메타데이터에 응답하여 제1 오디오 아이템의 압축을 적응시키도록 추가로 배열되는, 오디오 인코딩 장치.
- [0172] 12. 제1항 내지 제11항 중 어느 한 항에 있어서, 입력 프레젠테이션 메타데이터는 적어도 몇몇의 오디오 아이템들에 대한 우선순위 데이터를 포함하고, 인코더는 입력 프레젠테이션 메타데이터 내의 제1 오디오 아이템에 대한 우선순위 표시에 응답하여 제1 오디오 아이템에 대한 압축을 적응시키도록 배열되는, 오디오 인코딩 장치.
- [0173] 13. 제1항 내지 제12항 중 어느 한 항에 있어서, 오디오 인코더(205)는 인코딩이 입력 프레젠테이션 메타데이터에 응답하여 어떻게 적응되는지를 나타내는 인코딩 적응 데이터를 생성하도록, 그리고 인코딩 적응 데이터를 인코딩된 오디오 데이터 스트림에 포함시키도록 배열되는, 오디오 인코딩 장치.

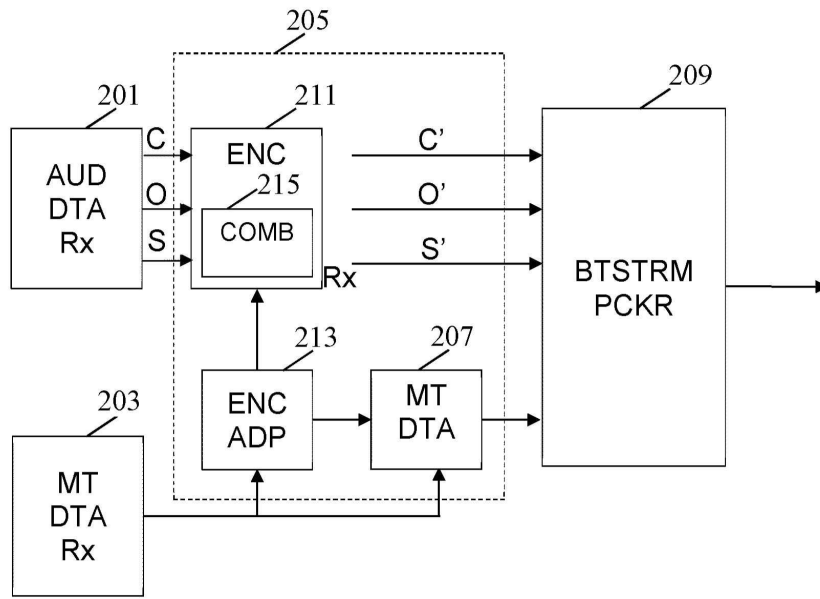
- [0174] 14. 오디오를 인코딩하는 방법으로서,
- [0175] 오디오 장면을 나타내는 복수의 오디오 아이템들을 수신하는 단계;
- [0176] 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계로서, 입력 프레젠테이션 메타데이터는 복수의 오디오 아이템들의 렌더링에 대한 프레젠테이션 제약들을 설명하는, 상기 복수의 오디오 아이템들에 대한 입력 프레젠테이션 메타데이터를 수신하는 단계;
- [0177] 복수의 오디오 아이템들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계로서, 인코딩은 입력 프레젠테이션 메타데이터에 응답하여 이루어지는, 상기 복수의 오디오 아이템들을 인코딩함으로써 오디오 장면에 대한 인코딩된 오디오 데이터를 생성하는 단계;
- [0178] 입력 프레젠테이션 메타데이터로부터 출력 프레젠테이션 메타데이터를 생성하는 단계; 및
- [0179] 인코딩된 오디오 데이터 및 출력 프레젠테이션 메타데이터를 포함하는 인코딩된 오디오 데이터 스트림을 생성하는 단계를 포함하는, 오디오를 인코딩하는 방법.
- [0180] 15. 컴퓨터 프로그램 제품으로서, 프로그램이 컴퓨터 상에서 실행될 때 제14항의 단계들 모두를 수행하도록 적응된 컴퓨터 프로그램 코드 수단을 포함하는, 컴퓨터 프로그램 제품.
- [0181] 보다 구체적으로, 본 발명은 첨부된 청구항들에 의해 한정된다.
- [0182] 본 발명이 몇몇 실시예들과 관련하여 설명되었지만, 본 명세서에 기재된 특정 형태로 제한되도록 의도되지 않는다. 오히려, 본 발명의 범위는 첨부된 청구항들에 의해서만 제한된다. 또한, 소정 특징이 특정 실시예들과 관련하여 설명되는 것처럼 보일 수 있지만, 당업자는 설명된 실시예들의 다양한 특징들이 본 발명에 따라 조합될 수 있음을 인식할 것이다. 청구항들에서, 용어 '포함하는'은 다른 요소들 또는 단계들의 존재를 배제하지 않는다.
- [0183] 또한, 개별적으로 열거되지만, 복수의 수단, 요소, 회로 또는 방법 단계는 예를 들어 단일 회로, 유닛 또는 프로세서에 의해 구현될 수 있다. 또한, 개개의 특징들이 상이한 청구항들에 포함될 수 있지만, 이들은 아마도 유리하게 조합될 수 있으며, 상이한 청구항들에의 포함은 특징들의 조합이 실현 가능하지 않고/않거나 유리하지 않다는 것을 암시하지는 않는다. 또한 하나의 카테고리의 청구항들에의 소정 특징의 포함은 이 카테고리로의 제한을 암시하는 것이 아니라, 오히려 그 특징이 적절한 바에 따라 다른 청구항 카테고리들에 동등하게 적용될 수 있음을 나타낸다. 또한, 청구항들에서의 특징들의 순서는 특징들이 작용되어야 하는 임의의 특정 순서를 암시하지는 않으며, 특히 방법 청구항에서의 개별 단계들의 순서는 단계들이 이 순서대로 수행되어야 함을 암시하지는 않는다. 오히려, 단계들은 임의의 적합한 순서로 수행될 수 있다. 또한, 단수형 언급은 복수를 배제하지 않는다. 이에 따라 "부정관사(a, an)", "제1", "제2" 등에 대한 언급은 복수를 배제하지 않는다. 청구항들에서의 참조 부호들은 단지 명료화 예로서 제공되며, 어떤 방식으로든 청구항들의 범위를 제한하는 것으로 해석되지 않아야 한다.

도면

도면1



도면2



103

도면3

