



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년01월12일
(11) 등록번호 10-2201713
(24) 등록일자 2021년01월06일

(51) 국제특허분류(Int. Cl.)
G10L 19/008 (2014.01)
(52) CPC특허분류
G10L 19/008 (2020.08)
(21) 출원번호 10-2020-7019184(분할)
(22) 출원일자(국제) 2013년07월19일
심사청구일자 2020년08월03일
(85) 번역문제출일자 2020년07월02일
(65) 공개번호 10-2020-0084918
(43) 공개일자 2020년07월13일
(62) 원출원 특허 10-2015-7001446
원출원일자(국제) 2013년07월19일
심사청구일자 2018년07월12일
(86) 국제출원번호 PCT/EP2013/065343
(87) 국제공개번호 WO 2014/013070
국제공개일자 2014년01월23일
(30) 우선권주장
12290239.8 2012년07월19일
유럽특허청(EPO)(EP)
(56) 선행기술조사문헌
KR1020140027954 A
(뒷면에 계속)

(73) 특허권자
돌비 인터네셔널 에이비
네덜란드 1101 씨엔 암스트레담 주이두스트 헤리
커베르그벡 1-35 3이 아폴로 빌딩
(72) 발명자
보엠, 요하네스
독일 30625 하노버 칼 비헤르트 알레 74 도이치
툼슨 오에이치쥐 리서치 앤드 이노베이션
잭스, 피터
독일 30625 하노버 칼 비헤르트 알레 74 도이치
툼슨 오에이치쥐 리서치 앤드 이노베이션
부에볼트, 올리버
독일 30625 하노버 칼 비헤르트 알레 74 도이치
툼슨 오에이치쥐 리서치 앤드 이노베이션
(74) 대리인
양영준, 백만기

전체 청구항 수 : 총 17 항

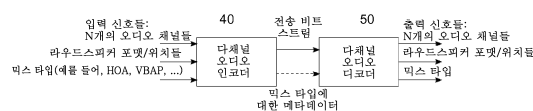
심사관 : 이남숙

(54) 발명의 명칭 다채널 오디오 신호들의 렌더링을 향상시키기 위한 방법 및 디바이스

(57) 요약

종래의 오디오 압축 기술들은 콘텐츠의 타입에 독립적인 표준화된 신호 변환을 수행한다. 다채널 신호들은 신호 컴포넌트들로 분해되고, 후속하여 양자화되고 인코딩된다. 이것은 씬(scene) 구성의 특성들, 구체적으로 예를 들어, 다채널 오디오 또는 HOA(Higher-Order Ambisonics) 콘텐츠에 대한 부족한 지식에 기인하여 바람직하지 않다. 선-프로세싱된 오디오 데이터를 인코딩하기 위한 향상된 방법은, 선-프로세싱된 오디오 데이터를 인코딩하는 단계, 및 특정 오디오 선-프로세싱을 나타내는 보조 데이터를 인코딩하는 단계를 포함한다. 인코딩된 오디오 데이터를 디코딩하기 위한 향상된 방법은, 인코딩된 오디오 데이터가 인코딩 전에 선-프로세싱되었다는 것을 결정하는 단계, 오디오 데이터를 디코딩하는 단계, 수신된 데이터로부터 선-프로세싱에 대한 정보를 추출하는 단계, 및 추출된 선-프로세싱 정보에 따라 디코딩된 오디오 데이터를 후-프로세싱하는 단계를 포함한다.

대표도 - 도2



(56) 선행기술조사문헌

US20040049379 A1*

US20120057715 A1*

ISO/IEC FDIS 23003-1:2006(E). Information technology - MPEG audio technologies Part 1: MPEG Surround. ISO/IEC JTC 1/SC 29/WG 11. 2006.07.21.*

Christof Faller. Parametric coding of spatial audio. PhD Thesis. ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE. 2004.*

*는 심사관에 의하여 인용된 문헌

명세서

청구범위

청구항 1

오디오 데이터를 인코딩하기 위한 방법으로서,

상기 오디오 데이터를 위해 HOA(Higher-Order Ambisonics) 포맷을 검출하는 단계;

iDST(inverse Discrete Spherical Harmonics Transform)에 기초하여 상기 HOA 포맷을 공통 HOA 포맷(common HOA format)으로 변환시키는 단계; 및

상기 공통 HOA 포맷의 계수들과, 상기 오디오 데이터에 관한 믹싱 정보 및 가상 또는 실제 라우드스피커 위치들에 관한 적어도 메타데이터를 나타내는 보조 데이터를 인코딩하는 단계

를 포함하고,

상기 믹싱 정보는 상기 HOA 포맷의 세부사항들을 포함하는, 방법.

청구항 2

제1항에 있어서,

상기 보조 데이터는, 상기 오디오 데이터가 HOA 콘텐츠로부터 도출되었다는 것과, HOA 콘텐츠 표현의 차수, 2D, 3D 또는 반구면 표현, 및 공간 샘플링 포인트들의 위치들 중 적어도 하나를 나타내는, 방법.

청구항 3

제1항에 있어서,

상기 보조 데이터는, 상기 오디오 데이터가 VBAP(vector-based amplitude panning)와, 라우드스피커들의 VBAP 튜플들(tupels) 또는 트리플들(triples)의 배치(assignment)를 이용하여 합성적으로 믹싱되었다는 것을 나타내는, 방법.

청구항 4

제1항에 있어서,

상기 보조 데이터는, 상기 오디오 데이터가 고정된, 이산 마이크로폰들과, 레코딩 세트 상의 하나 이상의 마이크로폰들의 하나 이상의 위치들과 방향들, 및 하나 이상의 종류의 마이크로폰들 중 적어도 하나를 이용하여 레코딩되었다는 것을 나타내는, 방법.

청구항 5

제1항에 있어서,

상기 HOA 포맷은, 복소수 값의 조화(complex-valued harmonics), 실수 값의 구면 조화(real-valued spherical harmonics) 및 정규화 체계(normalization scheme) 중 적어도 하나의 종류인, 방법.

청구항 6

인코딩된 오디오 데이터를 디코딩하기 위한 방법으로서,

인코딩된 오디오 데이터를 수신하는 단계; 및

상기 오디오 데이터를 디코딩하는 단계

를 포함하고,

상기 디코딩하는 단계는, 상기 오디오 데이터에 관한 믹싱 정보 및 가상 또는 실제 라우드스피커 위치들에 관한 적어도 메타데이터를 나타내는 보조 데이터를 디코딩하는 단계를 포함하고,

상기 믹싱 정보는 공통 HOA 포맷 및 추가 HOA 포맷의 세부사항들을 포함하고,

상기 디코딩하는 단계는, 상기 공통 HOA 포맷의 오디오 데이터를 상기 추가 HOA 포맷의 오디오 데이터로 변환시키는 단계를 더 포함하는, 방법.

청구항 7

제6항에 있어서,

상기 변환시키는 단계는, 상기 오디오 데이터가 제1 HOA 포맷을 갖는다는 지시자(indicator)에 기초하는 DSHT(Discrete Spherical Harmonics Transform)에 기초하는, 방법.

청구항 8

제6항에 있어서,

상기 적어도 메타데이터는, HOA 콘텐츠 표현의 차수, 2D, 3D 또는 반구면 표현, 및 공간 샘플링 포인트들의 위치들 중 적어도 하나에 관련되는, 방법.

청구항 9

제6항에 있어서,

상기 적어도 메타데이터는, 상기 오디오 데이터가 VBAP와, 라우드스피커들의 VBAP 튜플들 또는 트리플들의 배치에 기초하여 믹싱되었다는 것을 나타내는, 방법.

청구항 10

제6항에 있어서,

상기 적어도 메타데이터는, 상기 오디오 데이터가 고정된, 이산 마이크로폰들과, 하나 이상의 마이크로폰들의 적어도 위치와 적어도 방향, 및 적어도 마이크로폰의 종류 중 적어도 하나를 이용하여 레코딩되었다는 것을 나타내는, 방법.

청구항 11

제6항에 있어서,

상기 추가 HOA 포맷은, 복소수 값의 조화, 실수 값의 구면 조화 및 정규화 체계 중 적어도 하나의 종류인, 방법.

청구항 12

인코딩된 오디오 데이터를 디코딩하기 위한 장치로서,

상기 인코딩된 오디오 데이터가 인코딩 이전에 선-프로세싱되었다는 것을 결정하기 위한 분석기;

상기 오디오 데이터를 디코딩하기 위한 제1 디코더;

상기 선-프로세싱에 관한 정보를 추출하기 위한 데이터 스트림 파서 및 추출 유닛 - 상기 정보는 상기 오디오 데이터에 관한 믹싱 정보 및 가상 또는 실제 라우드스피커 위치들에 대한 적어도 메타데이터를 포함하고, 상기 믹싱 정보는 공통 HOA 포맷 및 추가 HOA 포맷의 세부사항들을 포함함 -; 및

상기 추출된 선-프로세싱 정보에 따라 상기 디코딩된 오디오 데이터를 후-프로세싱하기 위한 프로세싱 유닛을 포함하고,

상기 디코더는 상기 공통 HOA 포맷의 오디오 데이터를 상기 추가 HOA 포맷의 오디오 데이터로 변환시키도록 더 구성되는, 장치.

청구항 13

제12항에 있어서,

상기 디코더는 상기 오디오 데이터가 상기 추가 HOA 포맷을 갖는다는 지시자에 기초하는 DSHT에 기초하여 오디오 데이터를 변환시키도록 더 구성되는 장치.

청구항 14

제12항에 있어서,

상기 적어도 메타데이터는, HOA 콘텐츠 표현의 차수, 2D, 3D 또는 반구면 표현, 및 공간 샘플링 포인트들의 위치들 중 적어도 하나에 관련되는, 장치.

청구항 15

제12항에 있어서,

상기 적어도 메타데이터는, 상기 오디오 데이터가 VBAP와, 라우드스피커들의 VBAP 튜플들 또는 트리플들의 배치에 기초하여 믹싱되었다는 것을 나타내는, 장치.

청구항 16

제12항에 있어서,

상기 적어도 메타데이터는, 상기 오디오 데이터가 고정된, 이산 마이크로폰들과, 하나 이상의 마이크로폰들의 적어도 위치와 적어도 방향, 및 적어도 마이크로폰의 종류 중 적어도 하나를 이용하여 레코딩되었다는 것을 나타내는, 장치.

청구항 17

제12항에 있어서,

상기 추가 HOA 포맷은, 복소수 값의 조화, 실수 값의 구면 조화 및 정규화 체계 중 적어도 하나의 종류인, 장치.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 압축 분야에 관한 것으로, 특히 예를 들어 HOA(Higher Order Ambisonics)와 같은 다채널 오디오 신호들 및 사운드-필드-지향 오디오 씬들(sound-field-oriented audio scenes)의 압축에 관한 것이다.

배경 기술

[0002] 현재 다채널 오디오 신호들을 위한 압축 체계는 입력 오디오 자료가 어떻게 생성되었는지 또는 믹싱되었는지를 명시적으로 설명하지 않는다. 따라서, 알려진 오디오 압축 기술들은 압축해야 하는 콘텐츠의 원본/믹싱(origin/mixing) 타입을 알지 못한다. 알려진 접근법들에서, 다채널 신호가 신호 컴포넌트들로 분해되고, 후속하여 양자화되고 인코딩됨으로써, "블라인드(blind)" 신호 변환이 수행된다. 그러한 접근법들의 단점은, 상기 언급된 신호 분해의 연산이 계산적으로 요구되고, 주어진 오디오 씬의 세그먼트에 가장 적절하고 가장 효율적인 신호 분해를 찾기 어렵고 오류가 발생하기 쉽다는 것이다.

발명의 내용

해결하려는 과제

[0003] 본 발명은 다채널 오디오 렌더링을 향상시키기 위한 방법 및 디바이스에 관한 것이다.

[0004] 적어도 상기 언급된 단점들 중 일부는 씬 구성의 특성에 대한 부족한 사전 지식(knowledge)에 기인한다는 것을 발견하였다. 특히 공간 오디오 콘텐츠(spatial audio content), 예를 들어, 다채널 오디오 또는 HOA(Higher Order Ambisonics) 콘텐츠에 있어서, 압축 체계를 적용하는 데 이러한 사전 정보가 유용하다. 예를 들어, 압축 알고리즘에서의 일반적인 선-프로세싱 단계는 오디오 씬 분석으로서, 이 오디오 씬 분석에서는 원본 콘텐츠 또는 원본 콘텐츠 믹스로부터 방향성 오디오 소스들 또는 오디오 오브젝트들을 추출하는 것을 목표로 하고 있다. 그러한 방향성 오디오 소스들 또는 오디오 오브젝트들은 잔류 공간 오디오 콘텐츠는 별도로 코딩될 수 있다.

과제의 해결 수단

- [0005] 일 실시예에서, 선-프로세싱된 오디오 데이터를 인코딩하기 위한 방법은, 선-프로세싱된 오디오 데이터를 인코딩하는 단계, 및 특정 오디오 선-프로세싱을 나타내는 보조 데이터를 인코딩하는 단계를 포함한다.
- [0006] 일 실시예에서, 본 발명은 인코딩된 오디오 데이터를 디코딩하기 위한 방법에 관한 것으로, 인코딩된 오디오 데이터가 인코딩 전에 선-프로세싱되었다는 것을 결정하는 단계, 오디오 데이터를 디코딩하는 단계, 수신된 데이터로부터 선-프로세싱에 대한 정보를 추출하는 단계, 및 추출된 선-프로세싱 정보에 따라 디코딩된 오디오 데이터를 후-프로세싱하는 단계를 포함한다. 인코딩된 오디오 데이터가 인코딩 전에 선-프로세싱되었다는 것을 결정하는 단계는, 오디오 데이터의 분석에 의해 또는 수반되는 메타데이터의 분석에 의해 달성될 수 있다.
- [0007] 본 발명의 일 실시예에서, 선-프로세싱된 오디오 데이터를 인코딩하기 위한 인코더는, 선-프로세싱된 오디오 데이터를 인코딩하기 위한 제1 인코더, 및 특정 오디오 선-프로세싱을 나타내는 보조 데이터를 인코딩하기 위한 제2 인코더를 포함한다. 본 발명의 일 실시예에서, 인코딩된 오디오 데이터를 디코딩하기 위한 디코더는, 인코딩된 오디오 데이터가 인코딩 전에 선-프로세싱되었다는 것을 결정하기 위한 분석기, 오디오 데이터를 디코딩하기 위한 제1 디코더, 수신된 데이터로부터 선-프로세싱에 대한 정보를 추출하기 위한 데이터 스트림 파서(parser) 유닛 또는 데이터 스트림 추출 유닛, 및 추출된 선-프로세싱 정보에 따라 디코딩된 오디오 데이터를 후-프로세싱하기 위한 프로세싱 유닛을 포함한다.
- [0008] 본 발명의 일 실시예에서, 컴퓨터 판독가능 매체는 컴퓨터가 전술한 방법들 중 적어도 하나에 따른 방법을 컴퓨터가 수행하도록 야기하는 실행가능한 명령어들을 저장하고 있다.
- [0009] 본 발명의 일반적인 아이디어는 다음의 다채널 오디오 압축 시스템들의 확장들 중 적어도 하나에 기초한다.
- [0010] 일 실시예에 따라, 다채널 오디오 압축 및/또는 렌더링 시스템은, 다채널 오디오 신호 스트림(예를 들어, PCM 스트림들), 채널들 또는 대응하는 라우드스피커들의 관련된 공간 위치들, 및 다채널 오디오 신호 스트림에 적용되었던 믹싱 타입을 나타내는 메타데이터를 포함하는 인터페이스를 갖는다. 예를 들어, 믹싱 타입은 HOA 또는 VBAP 패닝, 특정 레코딩 기술들, 또는 균등 정보의 (이전의) 사용 또는 구성 및/또는 임의의 세부사항들을 나타낸다. 인터페이스는 신호 전송 체인에 대한 입력 인터페이스일 수 있다. HOA 콘텐츠의 경우에, 라우드스피커들의 공간 위치들은 가상 라우드스피커들의 위치들일 수 있다.
- [0011] 일 실시예에 따라, 다채널 압축 코덱의 비트 스트림은, 가상 또는 실제 라우드스피커 위치들 및 원본 믹싱 정보에 관한 전술한 메타데이터를 디코더와 후속 렌더링 알고리즘들로 전송하기 위한 시그널링 정보를 포함한다. 그렇게 함으로써, 디코딩 측에서 적용된 임의의 렌더링 기술들이 전송된 특정 콘텐츠의 인코딩 측 상의 특정 믹싱 특성들에 적용될 수 있다.
- [0012] 일 실시예에서, 메타데이터의 사용은 선택적이고, 스위치 온 또는 스위치 오프될 수 있다. 즉, 메타데이터를 이용하지 않고 단순 모드에서 오디오 콘텐츠가 디코딩되고 렌더링될 수 있지만, 단순 모드에서는 디코딩 및/또는 렌더링이 최적화되지 않을 것이다. 개선된 모드에서는, 메타데이터를 이용함으로써 최적화된 디코딩 및/또는 렌더링이 달성될 수 있다. 이 실시예에서, 디코더/렌더러는 2개의 모드들 사이에서 스위칭될 수 있다.

도면의 간단한 설명

- [0013] 본 발명의 바람직한 예시적인 실시예들은 첨부 도면들을 참조하여 설명된다.

도 1은 알려진 다채널 전송 시스템의 구조이다.

도 2는 본 발명의 일 실시예에 따른 다채널 전송 시스템의 구조이다.

도 3은 본 발명의 일 실시예에 따른 스마트 디코더이다.

도 4는 HOA 신호들에 대한 다채널 전송 시스템의 구조이다.

도 5는 DSHT의 공간 샘플링 포인트들이다.

도 6은 인코더 및 디코더 빌딩 블록들에 이용된 코드북에 대한 구면 샘플링 포인트들의 예들이다.

도 7은 특히 향상된 다채널 오디오 인코더의 예시적인 실시예이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 도 1은 다채널 오디오 코딩에 대한 알려진 접근법을 보여준다. 오디오 제작 스테이지(10)로부터의 오디오 데이터는 다채널 오디오 인코더(20)에서 인코딩되고, 다채널 오디오 디코더(30)에 전송되어 디코딩된다. 메타데이터는 명시적으로 전송되고(또는 그들의 정보가 암묵적으로 포함될 수 있음), 공간 오디오 구성에 관련된다. 그러한 사전 메타데이터는, 예를 들어, 특정 포맷들(예를 들어, 스테레오 또는 "5.1 서라운드 사운드(surround sound)"로 또한 알려진 ITU-R BS.775-1)의 형태로 또는 라우드스피커 위치들에 대한 표에 의한, 라우드스피커들의 공간 위치에 대한 정보에 제한된다. 특정 공간 오디오 믹스/레코딩이 어떻게 제작되었는지에 대한 정보가 다채널 오디오 인코더(20)에 전달되지 않기 때문에, 다채널 오디오 인코더(20) 내에서 신호를 압축하는 데 그러한 정보가 활용되거나 이용될 수 없다.
- [0015] 그러나, 다채널 공간 오디오 코더가, HOA(Higher Order Ambisonics) 포맷, 임의의 고정된 마이크로폰 셋업을 이용한 레코딩 및 임의의 특정 패닝 알고리즘들을 이용한 다채널 믹스로부터 파생되었던 콘텐츠 중 적어도 하나를 프로세싱하는 경우에, 콘텐츠의 원본 및 믹싱 타입 중 적어도 하나에 대한 지식이 특히 중요한 것으로 알려져 있는데, 이러한 경우에는 특정 믹싱 특성들이 압축 체계에 활용될 수 있기 때문이다. 또한 원본 다채널 오디오 콘텐츠는 추가 믹싱 정보 표시로부터 이득을 얻을 수 있다. 예를 들어, 인코딩 효율을 향상시키기 위한 VBAP(Vector-Based Amplitude Panning) 또는 그들의 임의의 세부사항들과 같은 이용된 패닝 방법을 표시하는 것이 바람직하다. 바람직하게, 후속 인코딩 단계들뿐만 아니라 오디오 씰 분석을 위한 신호 모델들이 이 정보에 따라 적응될 수 있다. 그 결과로, 왜곡률(rate-distortion) 성능과 연산 결과 모두에 대한 더 효율적인 압축 시스템을 얻게 된다.
- [0016] HOA 콘텐츠의 특정 경우에, 예를 들어, 복소수값 대 실수값의 구면 조화(spherical harmonics), 다중/상이한 정규화 체계들(normalization schemes) 등과 같은 다수의 상이한 종래의 기술들이 존재한다는 문제점이 있다. 상이하게 제작된 HOA 콘텐츠들 간의 비호환성을 피하기 위해, 공통 포맷을 정의하는 것이 유용하다. 이것은, DSHT(Discrete Spherical Harmonics Transform)과 같은 변환을 이용하여, 다채널 표현인 동일한 공간 표현에 대한 HOA 시간-도메인 계수들의 변환을 통해 달성될 수 있다. 공간 샘플링 위치들의 균일한 구면 분포로부터 DSHT가 생성되고, 이는 가상 라우드스피커 위치들과 동일하게 고려될 수 있다. DSHT에 대한 추가 정의들 및 세부사항들은 이하에 주어진다. HOA의 다른 정의를 이용하는 임의의 시스템은, 공간 도메인에서 정의된 이 공통 포맷으로부터 시스템 자체의 HOA 계수 표현을 얻을 수 있다. 이하에 더 상세하게 설명된 바와 같이, 상기 공통 포맷의 신호들의 압축은, 가상 라우드스피커 신호들이 원본 HOA 신호를 표현하는 사전 지식으로부터 상당한 이익을 얻게 된다.
- [0017] 또한, 이 믹싱 정보 등은 또한 디코더 또는 렌더러에 유용하다. 일 실시예에서, 믹싱 정보 등은 비트 스트림에 포함된다. 이용된 렌더링 알고리즘은, 예를 들어, HOA 또는 VBAP와 같은 원본 믹싱에 적응될 수 있는데, 더 나은 다운-믹스 또는 렌더링을 위해 유연한 라우드스피커 위치들을 가능하게 한다.
- [0018] 도 2는 본 발명의 일 실시예에 따른 다채널 오디오 전송 시스템의 확장을 도시한다. 오디오 콘텐츠의 제작 스테이지(10)에 적용되었던 믹싱 타입, 레코딩 타입, 편집 타입, 합성 타입 등 중 적어도 하나를 기술하는 메타데이터를 추가함으로써 확장이 이루어진다. 이 정보는 디코더 출력을 통해 전달되고 효율을 향상시키기 위해 다채널 압축 코덱(40, 50) 내에서 이용될 수 있다. 특정 공간 오디오 믹스/레코딩이 어떻게 제작되었는지에 대한 정보가 다채널 오디오 인코더(40)로 전달되어, 신호를 압축하는 데 활용되거나 이용될 수 있게 된다.
- [0019] 이 메타데이터 정보가 어떻게 이용될 수 있는지에 대한 한 예는, 입력 자료의 믹싱 타입에 의존하여 상이한 코딩 모드들이 다채널 코덱에 의해 활성화될 수 있다는 것이다. 예를 들어, 일 실시예에서, HOA 믹싱이 인코더 입력에서 표시되는 경우에, 이하에 설명된 바와 같이(수학식 3 내지 수학식 16 참조), HOA 특정 인코딩/디코딩 원리(HOA 모드)로 코딩 모드가 스위칭되고, 반면 입력 신호의 믹싱 타입이 HOA가 아니거나 또는 알려지지 않은 경우에, 상이한 (예를 들어, 더 전통적인) 다채널 코딩 기술이 이용된다. 일 실시예에서는, HOA 모드에서, HOA 특정 인코딩 프로세스가 시작되기 전에, DSHT가 원본 HOA 계수들을 다시 얻는 DSHT 블록을 이용하여 인코딩이 시작한다. 다른 실시예에서, 비교를 위해 DSHT와 다른 상이한 이산 변환이 이용된다.
- [0020] 도 3은 본 발명의 일 실시예에 따른 "스마트" 렌더링 시스템을 도시하는 것으로, 이러한 스마트 렌더링 시스템은 본 발명의 메타데이터를 이용하여, 디코더 단말기에 제공되는 M개의 라우드스피커들에 대한 디코딩된 N개의 채널들의 유연한 다운-믹스, 업-믹스 또는 리-믹스를 달성하게 한다. 효율적이고 고품질의 렌더링을 달성하기 위해, 믹싱, 레코딩 등의 타입에 대한 메타데이터는 복수의 모드들 중 하나를 선택하기 위해 활용될 수 있다. 다채널 인코더(50)는 입력 오디오 데이터에서 믹싱 타입에 대한 메타데이터에 따라 최적화된 인코딩을 이용하고, N개의 인코딩된 오디오 채널들 및 라우드 스피커 위치들에 대한 정보뿐만 아니라, 예를 들어, "믹스

타입" 정보를 인코딩하고 디코더(60)에 제공한다. M개의 오디오 채널들에 대한 출력 신호들을 생성하기 위해, (수신 측에서의) 디코더(60)는, 전송 측(즉, 인코더)에서는 알려지지 않은, 수신 측에서 이용가능한 라우드스피커들의 실제 라우드스피커 위치들을 이용한다. 일 실시예에서, N은 M과 상이하다. 일 실시예에서, N은 M과 동일하거나 또는 M과 상이하지만, 수신 측에서의 실제 라우드스피커 위치들은 인코더(50) 및 오디오 제작(10)에서 가정되었던 라우드스피커 위치들과 상이하다. 인코더(50) 또는 오디오 제작(10)은 예를 들어 표준화된 라우드스피커 위치들을 가정할 수 있다.

- [0021] 도 4는 본 발명이 어떻게 HOA 콘텐츠의 효율적인 전송을 위해 이용될 수 있는지를 도시한다. 입력 HOA 계수들은 역 DSHT(410; iDSHT)를 통해 공간 도메인으로 변환된다. 그 결과인 N개의 오디오 채널들, 그들의 (가상) 공간 위치들, 또한 표시(예를 들어, "HOA 믹싱된" 플래그와 같은 플래그)가 압축 인코더인 다채널 오디오 인코더(420)로 제공된다. 그에 따라 압축 인코더는 입력 신호가 HOA-도출되는 사전 지식을 이용할 수 있게 된다. 오디오 인코더(420) 및 오디오 디코더(430) 또는 오디오 렌더러 간의 인터페이스는 N개의 오디오 채널들, 그들의 (가상) 공간 위치들, 및 상기 표시를 포함한다. 역 프로세스는 디코딩 측에서 수행되는데, 즉, 디코딩(430) 이후에 콘텐츠를 인코딩하기 전에 적용되었던 관련된 동작들의 지식을 이용하는 DSHT(440)를 적용함으로써 HOA 표현이 복구될 수 있다. 이 지식은 본 발명에 따른 메타데이터의 형태에서 인터페이스를 통해 수신된다.
- [0022] 특히 본 발명의 범위 내에 있는 일부 (그러나 전체일 필요는 없음) 종류의 메타데이터는, 예를 들어,
- [0023] - 원본 콘텐츠가 HOA 콘텐츠로부터 도출되었고, 또한 이에 더하여,
- [0024] ○ HOA 표현의 차수,
- [0025] ○ 2D, 3D 또는 반구면 표현의 표시, 및
- [0026] ○ 공간 샘플링 포인트들의 위치들(적용 또는 고정) 중 적어도 하나로부터 파생되었다는 것을 나타내는 표시,
- [0027] - 원본 콘텐츠는 VBAP, 또한 이에 더하여 VBAP 튜플들(tupels) (쌍(pairs)) 또는 3중 라우드스피커들의 배치를 이용하여 합성적으로 믹싱되었다는 것을 나타내는 표시.
- [0028] - 원본 콘텐츠가 고정된, 이산 마이크로폰들을 이용하여 레코딩되었고, 또한 이에 더하여,
- [0029] 레코딩 세트 상의 하나 이상의 마이크로폰들의 하나 이상의 위치들 및 방향들, 및
- [0030] 예를 들어, 카디오이드(cardioid) 대 전방향(omnidirectional) 대 슈퍼-카디오이드(super-cardioid) 등과 같은 하나 이상의 종류의 마이크로폰들, 중 적어도 하나를 이용하여 레코딩되었다는 것을 나타내는 표시 중 적어도 하나일 것이다.
- [0031] 본 발명의 주요 장점은 적어도 다음과 같다.
- [0032] 입력 자료의 신호 특성들에 있어서, 더 나은 사전 지식을 통해 더 효율적인 압축 체계가 얻어진다. 인코더는 향상된 오디오 썬 분석을 위해 이러한 사전 지식을 활용할 수 있다(예를 들어, 믹싱된 콘텐츠의 소스 모델이 적용될 수 있음). 믹싱된 콘텐츠의 소스 모델에 대한 예는, 오디오 제작 스테이지(10)에서 신호 소스가 수정되고, 편집되고 또는 합성되었던 경우이다. 그러한 오디오 제작 스테이지(10)는 일반적으로 다채널 오디오 신호를 생성하는 데 이용되고, 일반적으로 다채널 오디오 인코더 블록(20) 전에 위치된다. 그러한 오디오 제작 스테이지(10)는 또한 도 2에서 신규 인코딩 블록(40) 전일 것으로 가정된다(그러나 도시되지 않음). 통상적으로, 편집 정보는 소실되어 인코더로 전달되지 않고, 그에 따라 활용될 수 없다. 본 발명은 이 정보가 보존될 수 있게 한다. 오디오 제작 스테이지(10)의 예들은 레코딩 및 믹싱, 합성 사운드 또는, 예를 들어, 라우드스피커 위치들에 합성하여 맵핑되는 다중 사운드 소스들과 같은 멀티-마이크로폰 정보를 포함한다.
- [0033] 본 발명의 다른 장점은, 유연한 라우드스피커 포지셔닝뿐만 아니라 특히 다수의 이용가능한 라우드스피커들이 다수의 이용가능한 채널들(소위 다운-믹스 및 업-믹스 시나리오들)과 상이한 나쁜 상황의 시나리오들에 있어서, 전송되고 디코딩된 콘텐츠의 렌더링이 상당히 향상될 수 있다는 것이다. 유연한 라우드스피커 포지셔닝은 라우드스피커 위치(들)에 따라 리-맵핑을 요구한다.
- [0034] 그밖에 다른 장점은, 고품질 렌더링을 위해 요구되는 중요한 데이터를 손실하지 않고, 채널-기반 오디오 전송 시스템들에서 오디오 데이터가 HOA와 같은 사운드 필드 관련 포맷으로 전송될 수 있다.
- [0035] 특히 공간 분해가 수행되는 경우에, 본 발명에 따른 메타데이터의 전송은 디코딩 측에서 최적화된 디코딩 및/또는 렌더링을 가능하게 한다. 다양한 수단에 의해 일반적인 공간 분해를 얻을 수 있으면서, 예를 들어,

KLT(Karhunen-Loeve Transform), 최적화된 분해(본 발명에 따른 메타데이터를 이용함)는 계산적으로 더 저렴하며, 동시에, 더 나은 품질의 다채널 출력 신호들을 제공한다(예를 들어, 신호 채널들이 렌더링 동안 라우드스피커 위치들에 더 용이하게 적응되고 맵핑될 수 있으며, 맵핑이 더 정확함). 이것은 특히, 렌더링 동안 믹싱(매트릭싱) 스테이지에서 채널들의 수가 수정되는 경우에(증가하거나 감소됨), 또는 하나 이상의 라우드스피커 위치들이 수정되는 경우에(특히 다채널의 각각의 채널들이 특정 라우드스피커 위치에 적응되는 경우) 바람직하다.

[0036] 다음에서, HOA(Higher Order Ambisonics) 및 DSHT(Discrete Spherical Harmonics Transform)이 설명된다.

[0037] 지각적 코더들을 이용하는 압축에 앞서, 예를 들어, DSHT(Discrete Spherical Harmonics Transform)에 의해 HOA 신호들은 공간 도메인으로 변환될 수 있다. 그러한 다채널 오디오 신호 표현들의 전송 또는 저장은 일반적으로 적절한 다채널 압축 기술들을 요구한다. 일반적으로, I개의 디코딩된 신호들 $\hat{x}_i(l), i = 1, \dots, I$ 에 J개의 신규 신호들 $\hat{y}_j(l), j = 1, \dots, J$ 로 최종 매트릭싱되기 전에, 채널 독립 지각적 디코딩이 수행된다. 용어 매트릭싱은 가중되는 방식으로 디코딩된 신호들 $\hat{x}_i(l)$ 를 추가하거나 믹싱하는 것을 의미한다. 다음과 같이 벡터에서 모든 신규 신호들 $\hat{y}_j(l), j = 1, \dots, J$ 뿐만 아니라 모든 신호들 $\hat{x}_i(l), i = 1, \dots, I$ 을 정리하면,

수학식 1

$$\hat{\mathbf{x}}(l) := [\hat{x}_1(l) \dots \hat{x}_I(l)]^T \quad (1a)$$

$$\hat{\mathbf{y}}(l) := [\hat{y}_1(l) \dots \hat{y}_J(l)]^T \quad (1b)$$

[0038]

[0039] 이고, 용어 "매트릭싱"은 수학적으로 매트릭스 오퍼레이션을 통해 $\hat{\mathbf{x}}(l)$ 로부터 $\hat{\mathbf{y}}(l)$ 를 얻게 된다는 사실로부터 유래한다.

수학식 2

$$\hat{\mathbf{y}}(l) = \mathbf{A} \hat{\mathbf{x}}(l)$$

[0040]

[0041] 여기서, A는 믹싱 가중치들을 포함하는 믹싱 매트릭스를 나타낸다. 본 명세서에서, 용어 "믹싱" 및 "매트릭싱"은 동의어로 이용된다. 믹싱/매트릭싱은 오디오 신호들을 임의의 특정 라우드스피커 셋업들을 위해 렌더링할 목적으로 이용된다.

[0042] 매트릭스가 의존하는 특정 개별 라우드스피커 셋업, 및 그에 따라 렌더링 동안 매트릭싱을 위해 이용되는 매트릭스는, 일반적으로 지각적 코딩 스테이지에 알려져 있지 않다.

[0043] 다음 섹션은 HOA(Higher Order Ambisonics)에 대한 간략한 소개를 제공하고, 프로세싱될 신호들(데이터 레이트 압축)을 정의한다.

[0044] HOA(Higher Order Ambisonics)는 사운드 소스들이 없음(free) 것으로 가정되는 소규모 관심 영역 내의 사운드 필드의 설명에 기초한다. 그러한 경우에, (구면좌표계에서의) 관심 영역 내의, 시간 t 및 위치 $\mathbf{x} = [r, \theta, \phi]^T$ 에서의 음압(sound pressure) p(t, x)의 시공간상의(spatiotemporal) 거동은, 동차 파동 방정식(homogeneous wave equation)에 의해 물리적으로 완전히 결정된다. 시간에 대한 음압의 푸리에 변환(Fourier transform)은 다음과 같이 보여질 수 있는데, 즉,

수학식 3

$$P(\omega, \mathbf{x}) = \mathcal{F}_t \{ p(t, \mathbf{x}) \}$$

[0045]

[0046] 이고, 여기서, ω 는 각주파수(그리고 $\mathcal{F}_t\{\}$ 는 $\int_{-\infty}^{\infty} p(t, \mathbf{x}) e^{-i\omega t} dt$ 에 대응함)를 나타내며, 다음에 따라 SHs(Spherical

Harmonics)의 수열로 확장될 수 있다.

수학식 4

$$P(k, c_s, \mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) j_n(kr) Y_n^m(\theta, \phi)$$

[0047]

수학식 4에, c_s 는 음속이고 $k = \frac{\omega}{c_s}$ 각파수(angular wave number)를 나타낸다. 또한, $j_n(\cdot)$ 은 제1종 구형 베셀 함수(the spherical Bessel functions of the first kind) 및 차수 n 이고, $Y_n^m(\cdot)$ 는 차수 n 및 수차(degree) m 에 대한 SH(Spherical Harmonics)를 나타낸다. 사운드 필드에 대한 완성된 정보는 사운드 필드 계수들 $A_n^m(k)$ 내에 실제로 포함된다. 일반적으로, SHs는 복소수 값 함수들이라는 것이 주목되어야 한다. 그러나, 적절한 선형 조합에 의해, 실수 값 함수들을 얻고 이러한 함수들에 대하여 확장하는 것이 가능하다.

[0048]

수학식 4에 음압 사운드 필드 설명과 관련하여, 소스 필드는 다음과 같이 정의될 수 있다.

[0049]

수학식 5

$$D(k, c_s, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n B_n^m(k) Y_n^m(\Omega)$$

[0050]

여기서, 소스 필드 또는 진폭 밀도[9] $D(kc_s, \Omega)$ 는 각파수(angular wave number) 및 각방향(angular direction) $\Omega = [\theta, \phi]^T$ 에 의존한다. 소스 필드는 먼-필드/가까운-필드, 이산/연속 소스들[1]을 포함할 수 있다. 사운드 필드 계수들 B_n^m 은 [1]에 의해 사운드 필드 계수들 A_n^m 에 관련된다.

[0051]

수학식 6

$$A_n^m = \begin{cases} 4\pi i^n B_n^m & \text{먼 필드} \\ -i k h_n^{(2)}(kr_s) B_n^m & \text{가까운 필드} \end{cases}$$

[0052]

여기서, $h_n^{(2)}$ 는 제2종 구면 한켈 함수(spherical Hankel function of the second kind)이고 r_s 는 원점으로부터의 소스 거리이다. 가까운 필드에 관하여, 양의 주파수들 및 제2종 구면 한켈 함수 $h_n^{(2)}$ 는 입력 파형(incoming waves)을 위해 이용된다(e^{-ikr} 과 관련됨).

[0053]

주파수 도메인에서 또는 시간 도메인에서, HOA 도메인에서의 신호들은 소스 필드 또는 사운드 필드 계수들의 역 푸리에 변환(inverse Fourier transform)으로서 표현될 수 있다. 다음 설명은 소스 필드 계수들의 시간 도메인 표현을 이용하는 것으로 가정할 것이다. 유한 수에 대하여,

[0054]

수학식 7

$$b_n^m = i\mathcal{F}_t \{ B_n^m \}$$

[0055]

수학식 5에서의 무한 수열은 $n = N$ 에서 절삭된다. 절삭은 공간 대역폭 제한에 대응한다. 계수들(또는 HOA 채널들)의 수는 다음에 의해 주어진다.

[0056]

수학식 8

[0057] $O_{3D} = (N + 1)^2$ for 3D

[0058] 또는 단지 2D 설명을 위해, $O_{2D} = 2N + 1$ 이 주어진다. 라우드스피커들에 의한 이후의 재생성을 위해, 계수들 b_n^m 은 단일 시간 샘플 m 의 오디오 정보를 포함한다. 계수들은 저장되거나 전송될 수 있고 그에 따라 데이터 레이트 압축이 가해진다. 계수들의 단일 시간 샘플 m 은 O_{3D} 개의 엘리먼트들을 갖는 벡터 $b(m)$ 으로 표현될 수 있다.

수학식 9

[0059] $b(m) := [b_0^0(m), b_1^{-1}(m), b_1^0(m), b_1^1(m), b_2^{-2}(m), \dots, b_N^N(m)]^T$

[0060] 그리고 매트릭스 B 에 의한 M 개의 시간 샘플들의 블록은 다음과 같다.

수학식 10

[0061] $B := [b(m_{START} + 1), b(m_{START} + 2), \dots, b(m_{START} + M)]$

[0062] 사운드 필드들의 2차원 표현들은 원형 조화들을 이용한 확장에 의해 유도될 수 있다. 이것은 상기 나타낸 일반 적인 설명에 대한 특정 경우에서와 같이, $\theta = \frac{\pi}{2}$ 의 고정된 경사도, 계수들의 상이한 가중 및 O_{2D} 개의 계수들($m = \pm n$)로 감소된 세트를 이용하여 제공될 수 있다. 따라서, 다음 고려사항들 전부는 또한 2D 표현들에 적용할 수 있고, 구(sphere)라는 용어는 원(circle)이라는 용어로 대체될 필요가 있게 된다.

[0063] 다음은 HOA 계수 도메인에서 공간 도메인, 채널 기반 도메인으로의 변환을 설명하고, 역으로 공간 도메인, 채널 기반 도메인에서 HOA 계수 도메인으로의 변환을 설명한다. 단위 구 상의 1개의 이산 공간 샘플 위치들 $\Omega_1 = [\Theta_1, \Phi_1]^T$ 에 대하여, 수학식 5는 시간 도메인 HOA 계수들을 이용하여 다시 쓰여질 수 있다.

수학식 11

[0064] $d_{\Omega_l} := \sum_{n=0}^N \sum_{m=-n}^n b_n^m Y_n^m(\Omega_l)$

[0065] $L_{Sd} = (N + 1)^2$ 개의 구면 샘플 위치들 Ω_1 를 가정하면, 이것은 HOA 데이터 블록 B 에 대하여 벡터 표현으로 다시 쓰여질 수 있다.

수학식 12

[0066] $W = \Psi_1 B$,

[0067] 여기서, $W := [w(m_{START} + 1), w(m_{START} + 2), \dots, w(m_{START} + M)]$ 이고, $w(m) = [d_{\Omega_1}(m), \dots, d_{\Omega_{Lsd}}(m)]^T$ 는 L_{Sd} 개의 다채널 신호의 단일 시간 샘플을 나타내며, 매트릭스 $\Psi_1 = [y_1, \dots, y_{Lsd}]^H$ 이고, 여기서 벡터들

$\mathbf{y}_l = [Y_0^0(\boldsymbol{\Omega}_l), Y_1^{-1}(\boldsymbol{\Omega}_l), \dots, Y_N^N(\boldsymbol{\Omega}_l)]^T$ 이다. 구면 샘플 위치들이 매우 균일하게 선택되는 경우에, 매트릭스 Ψ_f 는 다음과 같이 존재한다.

수학식 13

[0068] $\Psi_f \Psi_f^T = I,$

[0069] 여기서, I는 $0_{3D} \times 0_{3D}$ 단위행렬이다. 그 다음, 수학식 12에 대응하는 변환은 다음에 의해 정의될 수 있다.

수학식 14

[0070] $\mathbf{B} = \Psi_f \mathbf{W}.$

[0071] 수학식 14는 L_{sd} 개의 구면 신호들을 계수 도메인으로 변환하고 포워드 변환(forward transform)으로서 다시 쓰여질 수 있다.

수학식 15

[0072] $\mathbf{B} = DSHT\{\mathbf{W}\}$

[0073] 여기서 DSHT{ }는 이산 구면 조화 변환(Discrete Spherical Harmonics Transform)을 나타낸다. L_{sd} 개의 채널 기반 신호들을 형성하기 위해, 대응하는 역 변환은 0_{3D} 개의 계수 신호들을 공간 도메인으로 변환하고, 수학식 12는 다음과 같이 된다.

수학식 16

[0074] $\mathbf{W} = iDSHT\{\mathbf{B}\}$

[0075] HOA 계수들 0_{3D} (수학식 8 참조)의 수를 매칭하는 다수의 구면 위치들 L_{sd} 을 갖는 DSHT는 이하에 설명된다. 먼저, 디폴트 구면 샘플 그리드가 선택된다. M개의 시간 샘플들의 블록에 대하여, 구면 샘플 그리드가 회전되어,

수학식 17

[0076] $\sum_{l=1}^{L_{sd}} \sum_{j=1}^{L_{sd}} |\Sigma_{W_{sd,l,j}}| - \sum (\sigma_{sd_1}^2, \dots, \sigma_{sd_{L_{sd}}}^2)$

[0077] 상기 항의 로그가 최소화되는데, 여기서 $|\Sigma_{W_{sd,l,j}}|$ (행 인덱스 l 및 열 인덱스 j인 매트릭스)는 $\Sigma_{W_{sd}}$ 의 엘리먼트들의 절대값들이고, $\sigma_{sd_l}^2$ 은 $\Sigma_{W_{sd}}$ 의 대각선 엘리먼트들이다. 가시화된, 도 5에 도시된 바와 같이, 이것은 DSHT의 구면 샘플링 그리드에 대응한다.

[0078] DSHT의 적절한 구면 샘플 위치들 및 그러한 위치들을 유도하기 위한 절차들이 잘 알려져 있다. 샘플링 그리드의 예들이 도 6에 도시된다. 특히, 도 6은 인코더 및 디코더 빌딩 블록들(pE, pD)에서 이용된 코드북을 위한 구면 샘플링 위치들의 예들을 보여주는데, 즉, 도 6a에서 $L_{sd} = 4$, 도 6b에서 $L_{sd} = 9$, 도 6c에서 $L_{sd} = 16$, 그리고 도 6d에서 $L_{sd} = 25$ 이다. 그 중에서도, 선-정의된 공간 라우드스피커 구성들에 따라 렌더링하기 위해 그러한

코드북들이 이용될 수 있다.

[0079] 도 7은 도 4에 도시된 특히 향상된 다채널 오디오 인코더(420)의 예시적인 실시예를 보여준다. 다채널 오디오 인코더(420)는, (블록(410)의 역을 구하기 위해) 블록(410)의 역 DSHT에 역인 DSHT를 계산하는 DSHT 블록(421)을 포함한다. 블록(421)의 목적은 역 DSHT 블록(410)의 입력에 실질적으로 동일한 신호들(70)을 출력에 제공하는 것이다. 이 신호(70)의 프로세싱은 또한 최적화될 수 있다. 신호(70)는 MDCT 블록(422)에 제공되는 오디오 컴포넌트들뿐만 아니라, 하나 이상의 우세한 오디오 신호 컴포넌트들, 또는 더 정확히 말하면 하나 이상의 우세한 오디오 신호 컴포넌트들의 로케이션들을 표시하는 신호 부분들(71)을 포함한다. 이들은 적어도 하나의 가장 강한 소스 방향을 검출하고(424), iDSHT의 적응 회전을 위한 로테이션 파라미터들을 계산하기(425) 위해 이용된다. 일 실시예에서, 이것은 시간적으로 변하는데, 즉, 검출(424) 및 계산(425)은 정의된 이산 시간 단계들에서 연속적으로 재-적용된다. iDSHT에 대한 적응 회전 매트릭스가 계산되고 적응 iDSHT가 iDSHT 블록(423)에서 수행된다. 회전의 효과는, iDSHT(423)의 샘플링 그리드가 회전되어 측면들 중 한 측면(즉, 단일 공간 샘플 위치)이 가장 가장 소스 방향으로 매칭된다(이것은 시변일 수 있음)는 것이다. 이것은 더 높은 효율을 제공하고 그에 따라 iDSHT 블록(423)에 오디오 신호의 더 바람직한 인코딩을 제공하게 된다. MDCT 블록(422)은 오디오 프레임 세그먼트들의 시간적 오버랩핑을 보상하는 데 바람직하다. iDSHT 블록(423)은 인코딩된 오디오 신호(74)를 제공하고, 회전 파라미터 계산 블록(425)은 (적어도 부분적인) 선-프로세싱 정보(75)로서 회전 파라미터들을 제공한다. 게다가, 선-프로세싱 정보(75)는 다른 정보를 포함할 수 있다.

[0080] 또한, 본 발명은 다음 실시예들에 관한 것이다.

[0081] 일 실시예에서, 본 발명은 채널 기반 3D-오디오 표현을 전송하고 및/또는 저장하고 프로세싱하기 위한 방법에 관한 것으로, 채널 기반 오디오 정보에 따른 부가 정보(SI)를 송신하고/저장하는 단계를 포함하고, 부가 정보는 채널 기반 오디오 정보의 믹싱 타입 및 의도된 스피커 위치를 나타내고, 믹싱 타입은 이전 프로세싱 스테이지에서(예를 들어, 믹싱 스튜디오에서) 오디오 콘텐츠가 믹싱되었던 것에 따른 알고리즘을 나타내고, 스피커 위치들은 스피커들의 위치들(예를 들어, 믹싱 스튜디오에서의 이상적인 위치들) 또는 이전 프로세싱 스테이지의 가상 위치들을 나타낸다. 또한, 상기 데이터 구조 및 채널 기반 오디오 정보를 수신한 이후에 프로세싱 단계들은 믹싱 및 스피커 위치 정보를 이용한다.

[0082] 일 실시예에서, 본 발명은 채널 기반 3D 오디오 표현을 전송하고 및/또는 저장하고 프로세싱하기 위한 디바이스에 관한 것으로, 채널 기반 오디오 정보에 따른 부가 정보(SI)를 송신하기 위한 수단(또는 저장하기 위한 수단)을 포함하고, 부가 정보는 채널 기반 오디오 정보의 믹싱 타입 및 의도된 스피커 위치를 나타내고, 믹싱 타입은 이전 프로세싱 스테이지에서(예를 들어, 믹싱 스튜디오에서) 오디오 콘텐츠가 믹싱되었던 것에 따른 알고리즘을 시그널링하고, 스피커 위치들은 스피커들의 위치들(예를 들어, 믹싱 스튜디오에서의 이상적인 위치들) 또는 이전 프로세싱 스테이지의 가상 위치들을 나타낸다. 또한, 본 디바이스는, 상기 데이터 구조 및 채널 기반 오디오 정보를 수신한 이후에 믹싱 및 스피커 위치 정보를 이용하는 프로세서를 포함한다.

[0083] 일 실시예에서, 본 발명은, 이전에 채널 기반 표현으로 HOA 3D 오디오를 변환하는 데 이용되었던 이상적인 구면 샘플링 그리드에 관련된, HOA 콘텐츠, HOA 차수 및 가상 스피커 위치 정보를 믹싱 정보가 시그널링하는 3D 오디오 시스템에 관한 것이다. 전송된 채널 기반 오디오 정보 및 수반되는 부가 정보(SI)를 수신하고/판독한 이후에, 채널 기반 오디오를 HOA 포맷으로 재-인코딩하는 데 SI가 이용된다. 상기 구면 샘플링 위치들로부터의 모드-매트릭스 Ψ 를 계산함으로써, 그리고 그것을 채널 기반 콘텐츠와 승산하는 매트릭스(DSHT)를 계산함으로써 상기 재-인코딩이 실시된다.

[0084] 일 실시예에서, 시스템/방법은 상이한 HOA 포맷들에 대한 모호함을 피하기 위해 이용된다. 제작 측면에서 제1 HOA 포맷에서의 HOA 3D 오디오 콘텐츠는 제1 포맷에 관련된 iDSHT를 이용하여 관련된 채널 기반 3D 오디오 표현으로 변환되고 SI에 분배된다. 수신된 채널 기반 오디오 정보는 제2 포맷에 관련된 SI 및 DSHT를 이용하여 제2 HOA 포맷으로 변환된다. 본 시스템의 일 실시예에서, 제1 HOA 포맷은 복소수 값을 갖는 HOA 표현들을 이용하고, 제2 HOA 포맷은 실수 값을 갖는 HOA 표현들을 이용한다. 본 시스템의 일 실시예에서, 제2 HOA 포맷은 복소수 값을 갖는 HOA 표현을 이용하고, 제1 HOA 포맷은 실수 값을 갖는 HOA 표현을 이용한다.

[0085] 일 실시예에서, 본 발명은 3D 오디오 시스템에 관한 것으로, 믹싱 정보는 레이트 압축, 신호 향상 또는 렌더링에 이용된 신호로부터 방향성 3D 오디오 컴포넌트들 분리시키는 데 이용된다(오디오 오브젝트 추출). 일 실시예에서, 추가 단계들은, 이전에 HOA 3D 오디오를 채널 기반 표현으로 변환하는 데 이용되었던 HOA, HOA 차수 및 관련된 이상적인 구면 샘플링 그리드를 시그널링하고, HOA 표현들을 재저장하며, 블록 기반 공분산 방법들(covariance methods)을 이용하여 메인 신호 방향들을 결정함으로써 방향성 컴포넌트들을 추출한다. 상기 방향

들은 이러한 방향들로 방향성 신호들을 디코딩하는 HOA에 이용된다. 일 실시예에서, 추가 단계들은 VBAP(Vector Base Amplitude Panning) 및 관련된 스피커 위치 정보를 시그널링하는데, 여기서, 스피커 위치 정보는 스피커 트리플렛들(triplets)을 결정하는 데 이용되고, 공분산 방법은 상기 트리플렛 채널들 외부에서 상호 연관된 신호를 추출하는 데 이용된다.

[0086] 3D 오디오 시스템의 일 실시예에서, 신호 추출에 관련된 방향성 신호들 및 재저장된 신호들(HOA 신호들, VBAP 트리플렛들(쌍))로부터 잔류 신호들이 생성된다.

[0087] 일 실시예에서, 본 발명은 잔류 신호들의 데이터 레이트 압축을 수행하기 위한 시스템에 관한 것으로, HOA 잔류 신호의 차수를 감소시키는 단계, 감소된 차수 신호들 및 방향성 신호들을 압축하는 단계, 잔류 트리플렛 채널들을 모노 스트림으로 믹싱하고 관련된 상호 연관 정보를 제공하는 단계, 그리고 상기 정보 및 압축된 방향성 신호들과 함께 압축된 모노 신호들을 전송하는 단계에 의해 수행된다.

[0088] 데이터 레이트 압축을 수행하기 위한 시스템의 일 실시예에서, 시스템은 오디오를 라우드스피커들에 렌더링하기 위해 이용되는데, 채널 도메인에서 메인 신호 방향들 및 상호 비-연관된 잔류 신호들을 이용하여, 추출된 방향성 신호들이 라우드스피커들에 패닝된다.

[0089] 본 발명은 일반적으로 오디오 콘텐츠 믹싱 특성들의 신호전달을 가능하게 한다. 본 발명은, 오디오 디바이스들에서, 특히 오디오 인코딩 디바이스들, 오디오 믹싱 디바이스들 및 오디오 디코딩 디바이스들에서 이용될 수 있다.

[0090] DSHT로서 간략히 도시되었지만, DSHT와 다른 전송의 다른 타입들이 구성되거나 적용될 수 있다는 것이 주목되어야 하고, 본 발명의 의미 및 범위 내에 고려되는 모든 것들은 당업자에게 명백할 것이다. 또한, HOA 포맷이 상기 설명에 예시적으로 언급되었지만, 본 발명은 또한 앰비소닉과 다른 사운드필드 관련 포맷들의 다른 타입들로 이용될 수 있고, 본 발명의 의미 및 범위 내에 고려되는 모든 것들은 당업자에게 명백할 것이다.

[0091] 본 명세서의 바람직한 실시예들에 적용된 바와 같은 본 발명의 근본적인 새로운 특징들이 도시되고 설명되고 지시되었지만, 본 발명의 의미로부터 벗어나지 않고 설명된 장치 및 방법에서, 개시된 디바이스들의 형태 및 세부사항들에서, 그들의 동작에서, 당업자들에 의해 다양한 생략 및 보완 및 변화들이 이루어질 수 있다는 것이 이해될 것이다. 본 발명은 순전히 예의 방식으로 설명되었고, 본 발명의 범위로부터 벗어나지 않고 세부사항의 수정이 이루어질 수 있다는 것이 이해될 것이다. 동일한 결과들을 달성하기 위해 실질적으로 동일한 방법에서 실질적으로 동일한 기능을 수행하는 모든 엘리먼트들의 조합은 본 발명의 범위 내에 있는 것으로 명백하게 의도된다. 또한 설명된 실시예로부터 다른 것으로의 엘리먼트들의 보완들이 충분히 의도되고 고려된다.

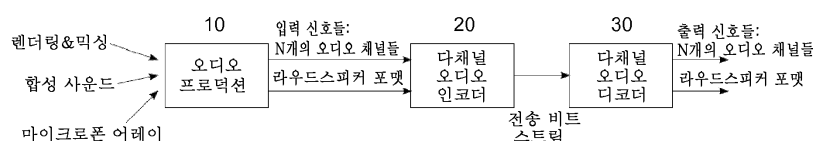
[0092] 참조

[0093] [1] T.D. Abhayapala "Generalized framework for spherical microphone arrays: Spatial and frequency decomposition", In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), (accepted) Vol. X, pp. , April 2008, Las Vegas, USA.

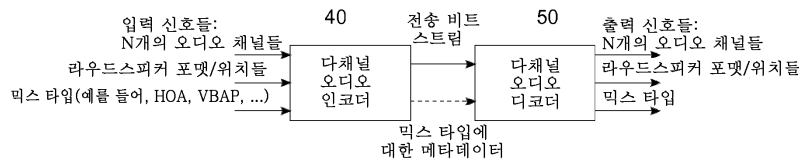
[0094] [2] James R. Driscoll and Dennis M. Healy Jr.: "Computing Fourier transforms and convolutions on the 2-sphere", Advances in Applied Mathematics, 15:202-250, 1994.

도면

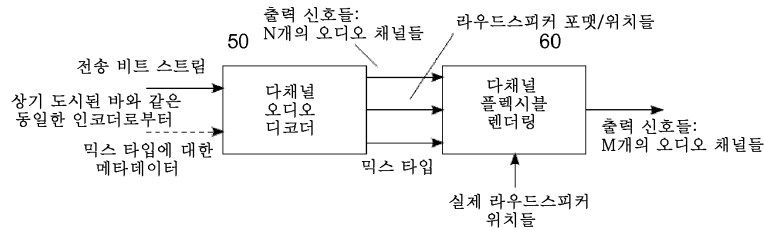
도면1



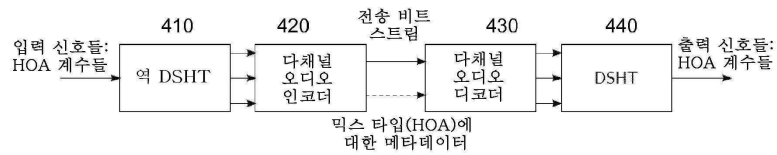
도면2



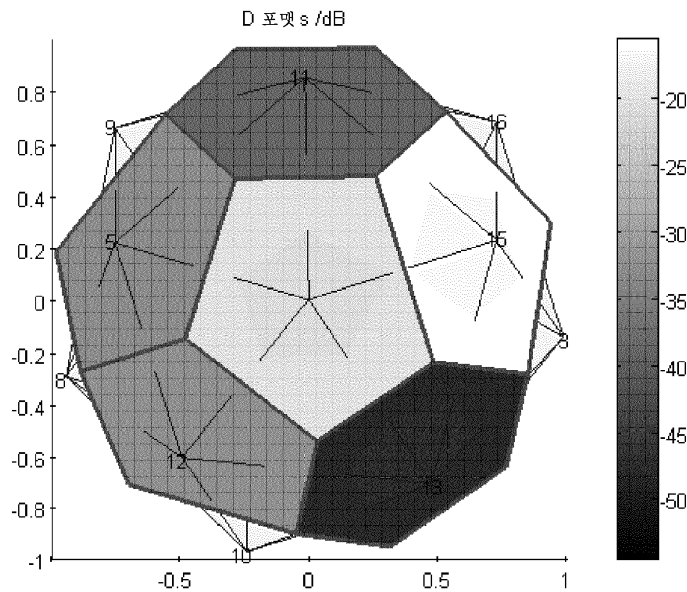
도면3



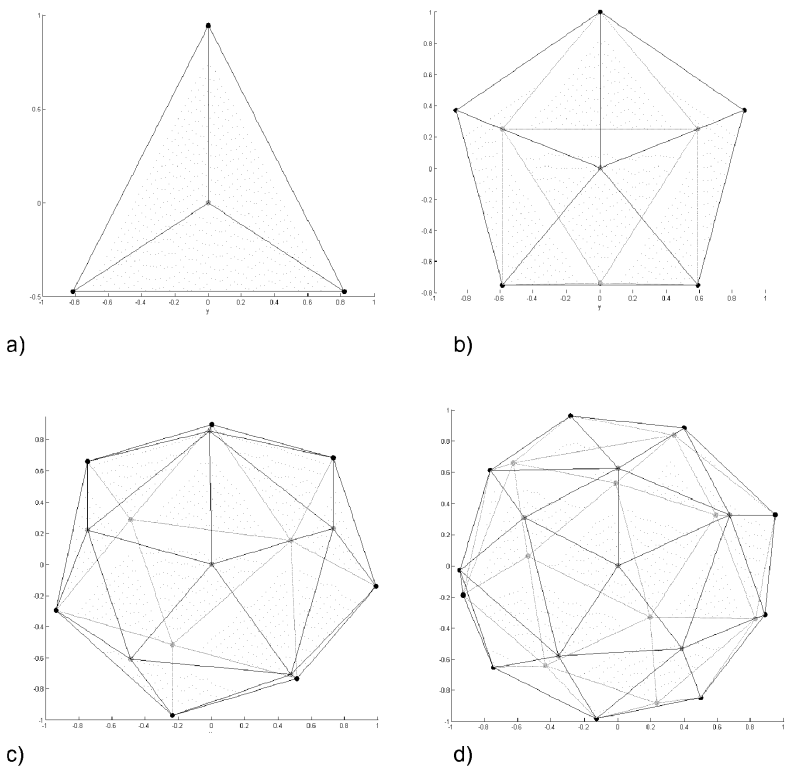
도면4



도면5



도면6



도면7

