



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0060464  
(43) 공개일자 2019년06월03일

(51) 국제특허분류(Int. Cl.)

H04S 7/00 (2006.01)

(52) CPC특허분류

H04S 7/303 (2013.01)

H04S 2420/01 (2013.01)

(21) 출원번호 10-2017-0158685

(22) 출원일자 2017년11월24일

심사청구일자 없음

(71) 출원인

주식회사 윌리스표준기술연구소

경기도 성남시 분당구 황새울로 216, 5층(수내동)

연세대학교 원주산학협력단

강원도 원주시 흥업면 연세대길 1

(72) 발명자

김동철

경기도 수원시 장안구 하북로46번길 17, 307동 506호

곽진삼

경기도 안양시 동안구 귀인로 213, 102동 1704호  
(뒷면에 계속)

(74) 대리인

임국일

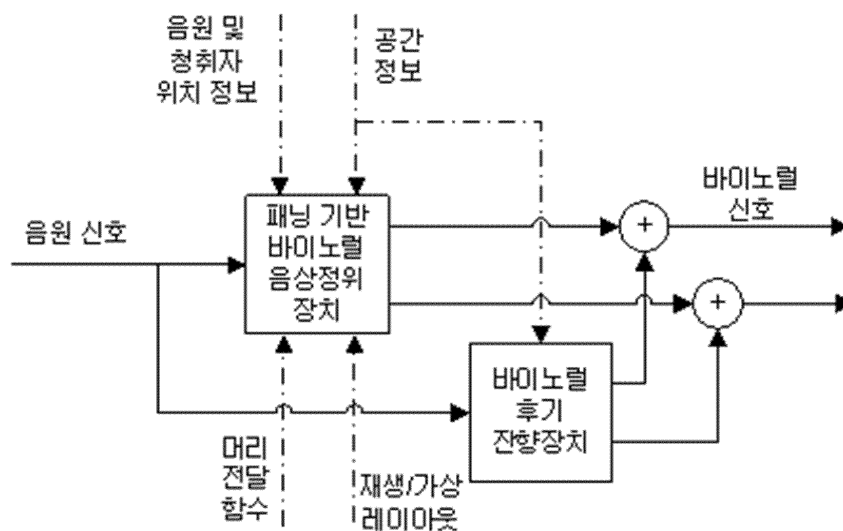
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 오디오 신호 처리 방법 및 장치

### (57) 요약

본 발명은 오디오 신호 처리 방법 및 장치에 관한 것으로, 더욱 상세하게는 오브젝트 신호와 채널 신호를 합성하고 이를 효과적으로 바이노럴 렌더링할 수 있는 오디오 신호 처리 방법 및 장치에 관한 것이다.

대표도 - 도5



(72) 발명자

**손주형**

경기도 의왕시 내손중앙로 11, 1114-302

**박영철**

강원도 원주시 시청로 264 (포스코더샵APT)  
105-401

**문현기**

서울시 서대문구 연희로 25길 4-5, 204호

이 발명을 지원한 국가연구개발사업

과제고유번호 10080057

부처명 산업통상자원부

연구관리전문기관 한국산업기술평가관리원

연구사업명 국가표준기술력향상사업

연구과제명 몰입형 가상현실 콘텐츠를 위한 MPEG- I 바이노럴 오디오 코덱 표준화

기 여 율 1/1

주관기관 (주)윌러스표준기술연구소

연구기간 2017.06.01 ~ 2020.12.31

---

## 명세서

### 청구범위

#### 청구항 1

오디오 신호 처리 방법 및 장치.

### 발명의 설명

#### 기술 분야

- [0001] 본 발명은 오디오 신호 처리 방법 및 장치에 관한 것으로, 더욱 상세하게는 오브젝트 신호와 채널 신호를 합성하고 이를 효과적으로 바이노럴 렌더링할 수 있는 오디오 신호 처리 방법 및 장치에 관한 것이다.

#### 배경 기술

- [0002] 3D 오디오란 기존의 서라운드 오디오에서 제공하는 수평면(2D) 상의 사운드 장면에 높이 방향에 해당하는 또 다른 축을 제공함으로써, 3차원 공간상에서 입장감 있는 사운드를 제공하기 위한 일련의 신호 처리, 전송, 부호화 및 재생기술 등을 통칭한다. 특히, 3D 오디오를 제공하기 위해서는 종래보다 많은 수의 스피커를 사용하거나 혹은 적은 수의 스피커를 사용하더라도 스피커가 존재하지 않는 가상의 위치에서 음상이 맺히도록 하는 렌더링 기술이 요구된다.
- [0003] 3D 오디오는 초고해상도 TV(UHDTV)에 대응되는 오디오 솔루션이 될 것으로 예상되며, 고품질 인포테인먼트 공간으로 진화하고 있는 차량에서의 사운드를 비롯하여 그밖에 극장 사운드, 개인용 3DTV, 태블릿, 스마트폰 및 클라우드 게임 등 다양한 분야에서 응용될 것으로 예상된다.
- [0004] 한편, 3D 오디오에 제공되는 음원의 형태로는 채널 기반의 신호와 오브젝트 기반의 신호가 존재할 수 있다. 이 뿐만 아니라, 채널 기반의 신호와 오브젝트 기반의 신호가 혼합된 형태의 음원이 존재할 수 있으며, 이를 통해 유저로 하여금 새로운 형태의 청취 경험을 제공할 수 있다.
- [0005] 한편, 오디오 신호 처리 장치에서 채널 기반 신호를 처리하기 위한 채널 렌더러와 오브젝트 기반 신호를 처리하기 위한 오브젝트 렌더러 간에는 성능 차이가 존재할 수 있다. 이를테면, 오디오 신호 처리 장치의 바이노럴 렌더링은 채널 기반 신호를 중심으로 구현될 수 있다. 이때, 오디오 신호 처리 장치의 입력으로 채널 기반 신호와 오브젝트 기반 신호가 혼합된 사운드 신(sound scene)이 수신될 경우, 바이노럴 렌더링을 통해 해당 사운드 신이 의도한 대로 재생되지 못할 수 있다. 따라서, 채널 렌더러와 오브젝트 렌더러 간의 성능 차이로 인해 발생할 수 있는 여러가지 문제점을 해결할 필요가 있다.

### 발명의 내용

#### 해결하려는 과제

- [0006] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 효과적으로 바이노럴 렌더링을 수행할 수 있는 장치 및 방법을 제공하고자 하는 목적을 가지고 있다.

#### 과제의 해결 수단

- [0007] 상기와 같은 과제를 해결하기 위해, 본 발명은 오디오 신호 처리 방법 및 오디오 신호 처리 장치를 제공한다.

#### 발명의 효과

- [0008] 본 발명의 실시예에 따르면, 바이노럴 렌더링을 효과적으로 수행할 수 있다.
- [0009] 또한, 파라미터화된 BRIR을 사용하여 낮은 연산량으로 고음질의 바이노럴 렌더링이 가능하도록 한다.
- [0010] 본 발명은 오디오 신호를 포함한 다양한 형태의 멀티미디어 신호의 필터링을 낮은 연산량으로 효율적으로 수행하는 방법을 제공한다.

## 도면의 간단한 설명

- [0011] 도 1은 본 발명의 실시예에 따른 오디오 인코더 및 오디오 디코더를 포함하는 전체 오디오 신호 처리 시스템을 나타낸 구성도.
- 도 2는 멀티채널 오디오 시스템의 일 실시예에 따른 멀티 채널 스피커의 배치를 나타낸 구성도.
- 도 3은 청취 공간상에서 3차원의 사운드 장면을 구성하는 각 사운드 오브젝트들의 위치를 개략적으로 나타낸 도면.
- 도 4는 본 발명의 실시예에 따른 오디오 신호 디코더를 나타낸 블록도.
- 도 5는 도 4의 또 다른 실시예로서 파라미터화된 BRIR 필터를 사용한 바이노럴 렌더링 과정을 나타낸 블록도
- 도 6은 도 5의 패닝 기반 바이노럴 음상정위 장치의 구성도

## 발명을 실시하기 위한 구체적인 내용

- [0012] 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어를 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도, 관례 또는 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한 특정 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 그 의미를 기재할 것이다. 따라서 본 명세서에서 사용되는 용어는, 단순한 용어의 명칭이 아닌 그 용어가 가진 실질적인 의미와 본 명세서의 전반에 걸친 내용을 토대로 해석되어야 함을 밝혀두고자 한다.
- [0013] 도 1은 본 발명의 실시예에 따른 오디오 인코더 및 오디오 디코더를 포함하는 전체 오디오 신호 처리 시스템을 나타낸 구성도이다.
- [0014] 도 1에 따르면 오디오 인코더(1100)는 입력 사운드 장면(sound scene)을 부호화하여 비트스트림을 생성한다. 오디오 디코더(1200)는 생성된 비트스트림을 수신할 수 있으며, 본 발명의 실시예에 따른 오디오 신호 처리 방법을 이용하여 해당 비트스트림을 디코딩 및 렌더링하여 출력 사운드 장면을 생성한다. 본 명세서에서 오디오 신호 처리 장치는 협의의 의미로는 오디오 디코더(1200)를 가리킬 수 있으나, 이에 한정하지 않으며 오디오 디코더(1200)에 포함된 세부 구성을 가리킬 수도 있고, 오디오 인코더(1100) 및 오디오 디코더(1200)를 포함하는 전체 오디오 신호 처리 시스템을 가리킬 수도 있다.
- [0015] 도 2는 멀티채널 오디오 시스템의 일 실시예에 따른 멀티채널 스피커의 배치(configuration)를 나타낸 구성도이다.
- [0016] 멀티채널 오디오 시스템에서는 임장감(presence)을 높이기 위해 복수의 스피커 채널이 사용될 수 있으며, 특히 3차원 공간상에서의 임장감을 제공하기 위해 너비, 깊이 및 높이 방향으로 복수의 스피커가 배치될 수 있다. 도 2는 일 실시예로써 22.2 채널의 스피커 배치를 도시하고 있으나, 본 발명은 특정 채널 수 또는 특정 스피커의 배치에 한정되지 않는다. 도 2를 참조하면, 22.2 채널의 스피커 셋은 상부 레이어(top layer), 중간 레이어(middle layer) 및 하부 레이어(bottom layer)의 3개의 레이어로 구성될 수 있다. TV 스크린의 위치를 전면이라고 할 때, 상부 레이어에는 전면에 3개, 중간 위치에 3개, 서라운드 위치에 3개가 배치되어 총 9개의 스피커가 배치될 수 있다. 또한, 중간 레이어에는 전면에 5개, 중간 위치에 2개, 서라운드 위치에 3개가 배치되어 총 10개의 스피커가 배치될 수 있다. 한편, 하부 레이어에는 전면에 3개의 스피커가 배치되고, 2개의 LFE 채널 스피커가 구비될 수 있다.
- [0017] 이와 같이 최대 수십 개의 채널에 이르는 멀티채널 신호를 전송 및 재생하기 위해서는 높은 연산량이 요구된다. 또한, 통신 환경 등을 고려할 때, 해당 신호에 대한 높은 압축률이 요구될 수 있다. 뿐만 아니라, 일반 가정에서는 22.2 채널과 같은 멀티채널 스피커 시스템을 구비하는 유저는 극히 드물고, 2 채널 또는 5.1 채널 셋업을 갖는 시스템이 구비되는 경우가 많다. 따라서, 모든 유저에게 공통적으로 전송되는 신호가 멀티채널을 각각 인코딩한 신호인 경우, 해당 멀티채널 신호를 다시 2 채널 또는 5.1 채널에 대응하도록 변환하는 과정이 필요하다. 이에 따라, 통신적인 비효율이 발생할 뿐만 아니라, 22.2 채널의 PCM(Pulse Code Modulation) 신호를 저장해야 하므로 메모리 관리에 있어서도 비효율적인 문제가 발생할 수 있다.
- [0018] 도 3은 청취 공간상에서 3차원의 사운드 장면을 구성하는 각 사운드 오브젝트들의 위치를 개략적으로 나타내고 있다.
- [0019] 도 3에 도시된 바와 같이, 청자(52)가 3D 오디오를 청취하는 청취 공간(50) 상에서 3차원의 사운드 장면을 구성

하는 각 사운드 오브젝트(51)들의 위치는 점 소스(point source) 형태로 다양한 위치에 분포될 수 있다. 뿐만 아니라, 사운드 장면에는 점 소스 이외에도 평면파(plain wave) 형태의 음원이나, 앰비언트(ambient) 음원 등이 포함될 수 있다. 이와 같이 3차원 공간상에 다양하게 분포 되어 있는 오브젝트 및 음원들을 청자(52)에게 명확하게 제공하기 위해서는 효율적인 렌더링 방법이 필요하다.

[0020] 도 4는 본 발명의 실시예에 따른 오디오 디코더를 나타낸 블록도이다. 본 발명의 오디오 디코더(1200)는 코어 디코더(10), 렌더링 유닛(20), 믹서(30), 및 포스트 프로세싱 유닛(40)을 포함한다.

[0021] 먼저, 코어 디코더(10)는 수신된 비트스트림을 복호화하여 렌더링 유닛(20)으로 전달한다. 이때, 코어 디코더(10)에서 출력되어 렌더링 유닛으로 전달되는 신호에는 라우드스피커(loudspeaker) 채널 신호(411), 오브젝트 신호(412), SAOC 채널 신호(414), HOA 신호(415) 및 오브젝트 메타데이터 비트스트림(413) 등이 포함될 수 있다. 코어 디코더(10)에는 인코더에서 부호화시에 사용된 코어 코덱이 사용될 수 있는데, 이를테면, MP3, AAC, AC3 또는 USAC(Unified Speech and Audio Coding) 기반의 코덱이 사용될 수 있다.

[0022] 한편, 수신된 비트스트림에는 코어 디코더(10)에서 복호화되는 신호가 채널 신호인지, 오브젝트 신호인지 또는 HOA 신호인지 등을 식별할 수 있는 식별자가 더 포함될 수 있다. 또한, 복호화되는 신호가 채널 신호(411)일 경우, 각 신호가 멀티채널 내의 어느 채널 (이를테면 left speaker 대응, top rear right speaker 대응 등)에 대응되는지를 식별할 수 있는 식별자가 비트스트림에 더 포함될 수 있다. 복호화되는 신호가 오브젝트 신호(412)일 경우, 오브젝트 메타데이터 비트스트림(413)을 복호화하여 획득되는 오브젝트 메타데이터 정보(425a, 425b) 등과 같이, 해당 신호가 재생 공간의 어느 위치에 재생되는지를 나타내는 정보가 추가로 획득될 수 있다.

[0023] 본 발명의 실시예에 따르면, 오디오 디코더는 유연한 렌더링(flexible rendering)을 수행하여 출력 오디오 신호의 품질을 높일 수 있다. 유연한 렌더링이란 실제 재생 환경의 라우드스피커 배치(재생 레이아웃) 또는 BRIR(Binaural Room Impulse Response) 필터 셋의 가상 스피커 배치(가상 레이아웃)에 기초하여, 복호화된 오디오 신호의 포맷을 변환하는 과정을 의미할 수 있다. 일반적으로, 실제 거실 환경에 배치된 스피커는 규격(standard) 권고안 대비 방향각과 거리 등이 모두 달라지게 된다. 스피커의 높이, 방향, 청자와의 거리 등이 규격 권고안에 따른 스피커 배치와 상이하게 됨에 따라, 변경된 스피커의 위치에서 원래 신호를 재생할 경우 이 상적인 3D 사운드 장면을 제공하기 어렵게 될 수 있다. 이와 같이 상이한 스피커 배치에서도 콘텐츠 제작자가 의도한 사운드 장면을 효과적으로 제공하기 위해서는, 오디오 신호를 변환하여 스피커들 간의 위치 차이에 따른 변화를 보정하는 유연한 렌더링이 필요하다.

[0024] 따라서, 렌더링 유닛(20)은 코어 디코더(10)에 의해 복호화 된 신호를 재생 레이아웃(reproduction layout) 정보 또는 가상 레이아웃(virtual layout) 정보를 이용하여 타겟 출력 신호로 렌더링한다. 재생 레이아웃 정보는 타겟 채널의 배치(configuration)를 나타내며, 재생 환경의 라우드스피커 레이아웃 정보로 표현될 수 있다. 또한, 가상 레이아웃 정보는 바이노럴 렌더러(200)에서 사용되는 BRIR(Binaural Room Impulse Response) 필터 셋(set)에 기초하여 획득될 수 있는데, 가상 레이아웃에 대응하는 위치 셋(set of positions)은 BRIR 필터 셋에 대응하는 위치 셋의 서브셋(subset)으로 이루어 질 수 있다. 이때, 상기 가상 레이아웃의 위치 셋은 각 타겟 채널들의 위치 정보를 나타낸다. 렌더링 유닛(20)은 포맷 컨버터(22), 오브젝트 렌더러(24), OAM 디코더(25), SAOC 디코더(26) 및 HOA 디코더(28)를 포함할 수 있다. 렌더링 유닛(20)은 복호화 된 신호의 타입에 따라 상기 구성 중 적어도 하나를 이용하여 렌더링을 수행한다.

[0025] 포맷 컨버터(22)는 채널 렌더러로도 지칭될 수 있으며, 전송된 채널 신호(411)를 출력 스피커 채널 신호로 변환한다. 즉, 포맷 컨버터(22)는 전송된 채널 배치(configuration)와 재생될 스피커 채널 배치 간의 변환을 수행한다. 만약, 출력 스피커 채널의 개수(이를테면, 5.1 채널)가 전송된 채널의 개수(이를테면, 22.2 채널)보다 적거나, 전송된 채널 배치와 재생될 채널 배치가 다를 경우, 포맷 컨버터(22)는 채널 신호(411)에 대한 다운믹스 또는 변환을 수행한다. 본 발명의 실시예에 따르면, 오디오 디코더는 입력 채널 신호와 출력 스피커 채널 신호간의 조합을 이용하여 최적의 다운믹스 매트릭스를 생성하고, 상기 매트릭스를 이용하여 다운믹스를 수행할 수 있다. 또한, 포맷 컨버터(22)가 처리하는 채널 신호(411)에는 사전-렌더링된 오브젝트 신호가 포함될 수 있다. 일 실시예에 따르면, 오디오 신호의 부호화 전에 적어도 하나의 오브젝트 신호가 사전-렌더링되어 채널 신호에 믹스(mix)될 수 있다. 이와 같이 믹스 된 오브젝트 신호는 채널 신호와 함께 포맷 컨버터(22)에 의해 출력 스피커 채널 신호로 변환될 수 있다.

[0026] 오브젝트 렌더러(24) 및 SAOC 디코더(26)는 오브젝트 기반의 오디오 신호에 대한 렌더링을 수행한다. 오브젝트 기반의 오디오 신호에는 개별 오브젝트 웨이브폼과 파라메트릭 오브젝트 웨이브폼이 포함될 수 있다. 개별 오브젝트 웨이브폼의 경우, 각 오브젝트 신호들은 모노포닉(monophonic) 웨이브폼으로 인코더에 제공되며, 인코더

는 단일 채널 엘리먼트들(Single Channel Elements, SCEs)을 이용하여 각 오브젝트 신호들을 전송한다. 파라메트릭 오브젝트 웨이브폼의 경우, 복수의 오브젝트 신호들이 적어도 하나의 채널 신호로 다운믹스되며, 각 오브젝트의 특징과 이들 간의 관계가 SAOC(Spatial Audio Object Coding) 파라미터로 표현된다. 오브젝트 신호들은 다운믹스되어 코어 코덱으로 부호화되며, 이때 생성되는 파라메트릭 정보가 함께 디코더로 전송된다.

[0027] 한편, 개별 오브젝트 웨이브폼 또는 파라메트릭 오브젝트 웨이브폼이 오디오 디코더로 전송될 때, 이에 대응하는 압축된 오브젝트 메타데이터가 함께 전송될 수 있다. 오브젝트 메타데이터는 오브젝트 속성을 시간과 공간 단위로 양자화하여 3차원 공간에서의 각 오브젝트의 위치 및 이득값을 지정한다. 렌더링 유닛(20)의 OAM 디코더(25)는 압축된 오브젝트 메타데이터 비트스트림(413)을 수신하고, 이를 복호화하여 오브젝트 렌더러(24) 및/또는 SAOC 디코더(26)로 전달한다.

[0028] 오브젝트 렌더러(24)는 오브젝트 메타데이터 정보(425a)를 이용하여 각 오브젝트 신호(412)를 주어진 재생 포맷에 따라 렌더링한다. 이때, 각 오브젝트 신호(412)는 오브젝트 메타데이터 정보(425a)에 기초하여 특정 출력 채널들로 렌더링될 수 있다. SAOC 디코더(26)는 SAOC 채널 신호(414)와 파라메트릭 정보로부터 오브젝트/채널 신호를 복원한다. 또한, 상기 SAOC 디코더(26)는 재생 레이아웃 정보와 오브젝트 메타데이터 정보(425b)에 기초하여 출력 오디오 신호를 생성할 수 있다. 즉, SAOC 디코더(26)는 SAOC 채널 신호(414)를 이용하여 복호화된 오브젝트 신호를 생성하고, 이를 타겟 출력 신호로 매핑하는 렌더링을 수행한다. 이와 같이 오브젝트 렌더러(24) 및 SAOC 디코더(26)는 오브젝트 신호를 채널 신호로 렌더링할 수 있다.

[0029] HOA 디코더(28)는 HOA(Higher Order Ambisonics) 신호(415) 및 HOA 부가 정보를 수신하고, 이를 복호화한다. HOA 디코더(28)는 채널 신호나 오브젝트 신호를 별도의 수학적식으로 모델링하여 사운드 장면을 생성한다. 생성된 사운드 장면에서 스피커가 있는 공간상의 위치를 선택하면, 스피커 채널 신호로 렌더링이 수행될 수 있다.

[0030] 한편, 도 4에는 도시되지 않았지만, 렌더링 유닛(20)의 각 구성요소로 오디오 신호가 전달될 때, 전처리 과정으로서 동적 범위 제어(Dynamic Range Control, DRC)가 수행될 수 있다. DRC는 재생되는 오디오 신호의 동적 범위를 일정 레벨로 제한하는 것으로, 기 설정된 스레숄드(threshold) 보다 작은 소리는 더 크게, 기 설정된 스레숄드 보다 큰 소리는 더 작게 조정한다.

[0031] 렌더링 유닛(20)에 의해 처리된 채널 기반의 오디오 신호 및 오브젝트 기반의 오디오 신호는 믹서(30)로 전달된다. 믹서(30)는 렌더링 유닛(20)의 각 서브 유닛에서 렌더링된 부분 신호들을 믹싱하여 믹서 출력 신호를 생성한다. 만약 부분 신호들이 재생/가상 레이아웃 상의 동일한 위치에 매칭되는 신호일 경우에는 서로 더해지며, 동일하지 않은 위치에 매칭되는 신호일 경우에는 각각 별개의 위치에 대응되는 출력 신호로 믹싱된다. 믹서(30)는 서로 더해지는 부분 신호들 간에 상쇄 간섭이 발생하는지 여부를 판별하고, 이를 방지하기 위한 추가적인 프로세스를 더 수행할 수 있다. 또한, 믹서(30)는 채널 기반의 웨이브폼과 렌더링된 오브젝트 웨이브폼의 딜레이(delay)를 조정하고, 이를 샘플 단위로 합산한다. 이와 같이, 믹서(30)에 의해 합산된 오디오 신호는 포스트 프로세싱 유닛(40)으로 전달된다.

[0032] 포스트 프로세싱 유닛(40)은 스피커 렌더러(100)와 바이노럴 렌더러(200)를 포함한다. 스피커 렌더러(100)는 믹서(30)로부터 전달된 멀티채널 및/또는 멀티오브젝트 오디오 신호를 출력하기 위한 포스트 프로세싱을 수행한다. 이러한 포스트 프로세싱에는 동적 범위 제어(DRC), 음량 정규화(Loudness Normalization, LN) 및 피크 제한(Peak Limiter, PL) 등이 포함될 수 있다. 스피커 렌더러(100)의 출력 신호는 멀티채널 오디오 시스템의 라우드스피커로 전달되어 출력될 수 있다.

[0033] 바이노럴 렌더러(200)는 멀티채널 및/또는 멀티오브젝트 오디오 신호의 바이노럴 다운믹스 신호를 생성한다. 바이노럴 다운믹스 신호는 각 입력 채널/오브젝트 신호가 3차원상에 위치한 가상의 음원에 의해 표현되도록 하는 2채널의 오디오 신호이다. 바이노럴 렌더러(200)는 스피커 렌더러(100)에 공급되는 오디오 신호를 입력 신호로서 수신할 수 있다. 바이노럴 렌더링은 BRIR(Binaural Room Impulse Response) 필터를 기초로 수행되며, 시간 도메인 또는 QMF 도메인 상에서 수행될 수 있다. 실시예에 따르면, 바이노럴 렌더링의 후처리 과정으로서 전술한 동적 범위 제어(DRC), 음량 정규화(LN) 및 피크 제한(PL) 등이 추가로 수행될 수 있다. 바이노럴 렌더러(200)의 출력 신호는 헤드폰, 이어폰 등과 같은 2채널 오디오 출력 장치로 전달되어 출력될 수 있다.

[0034] 도 5는 도 4의 기능을 확장한 또 다른 실시예이다. 본 실시예는 BRIR 필터 대신 파라미터화된 직접음과 초기 반사음, 후기 잔향을 사용하여 바이노럴 렌더링을 수행한다.

[0035] BRIR은 일반적으로 직접음과 초기반사음, 후기잔향으로 나뉜다. BRIR의 직접음은 음원과 청취자의 상대적인 거리차이로 인한 시간지연과 방위각 및 고도각의 차이에 해당하는 HRTF(Head Related Transfer Function) 필터로



파라미터화 할 수 있다. BRIR의 초기 반사음은 Image source method, ray tracing 등의 geometrical room acoustic modeling을 사용하여 그 위치와 청취자까지의 음원 전파 경로가 추정될 수 있다. 음원 전파 경로의 주파수 응답은 공간 정보를 사용하여 추정할 수 있다. 초기반사음 또한 시간지연과 HRTF 및 음원 전파 경로에 대한 주파수 응답 필터로 파라미터화 할 수 있다.

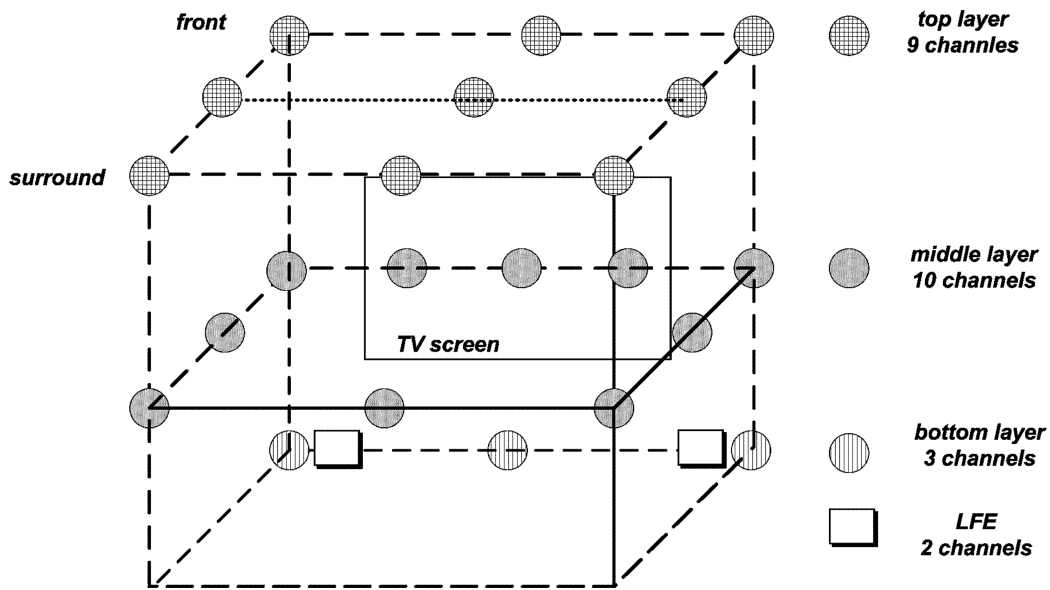
- [0036] 도 5의 패닝 기반 바이노럴 음상정위 장치는 음원 신호의 직접음과 초기 반사음을 생성하고 바이노럴 후기 잔향 장치는 음원 신호의 바이노럴 후기 잔향음을 생성한다. 공간 정보는 파라미터화한 BRIR을 생성하기 위한 정보로서 음원 신호와 청취자가 위치한 공간의 음향학적인 특성을 포함한다.
- [0037] 도 6은 도 5의 패닝 기반 바이노럴 음상정위 장치의 상세 블록도를 나타낸다. 시간 지연 계산부는 직접음과 초기 반사음의 시간 지연을 계산하고 시간 지연부는 계산된 시간 지연을 음원 신호에 적용한다. 필터 계수 계산부는 직접음과 초기 반사음의 주파수 응답에 해당하는 필터 계수를 생성한다. 필터 연산부는 시간 지연된 음원 신호에 생성된 필터 계수를 적용한다.
- [0038] 패닝 계수 계산부는 직접음 혹은 초기 반사음과 청취자의 상대적인 위치와 재생/가상 레이아웃을 사용하여 해당 재생/가상 레이아웃에 적합한 멀티채널 신호로 변환하기 위한 패닝 계수를 생성한다. 패닝 연산부는 시간 지연 및 주파수 응답이 적용된 음원 신호에 생성된 패닝 계수를 적용한다. 최종적으로 음원 신호의 직접음과 초기 반사음을 멀티채널 음원으로 구현된다.
- [0039] 제안한 구조는 렌더링 시스템의 연산량에 따라 재생/가상 레이아웃의 공간 해상도를 조절할 수 있으며, 청각의 공간 해상도가 높은 정면에는 높은 공간 해상도로 채널을 배치하고, 이외의 부분에는 낮은 공간 해상도로 채널을 배치함으로써 렌더링 품질을 최대한 유지하면서 연산량을 줄일 수 있다. 또한 정확한 음상정위가 필요한 객체의 경우 높은 공간 해상도의 HRTF 세트를 사용하여 바이노럴 렌더링하고, 나머지 객체와 반사음은 낮은 공간 해상도의 HRTF 세트와 패닝 기반 바이노럴 렌더링 기법을 적용하는 방법이 사용 가능하다.
- [0040] MPEG-H 3D 오디오 표준은 시작단계부터 사용자의 움직임이 없는 상황을 가정하여 만들어진 표준이다. 청취자의 움직임 및 오디오 장면과 청취자간의 상호작용으로 인하여 변화하는 BRIR을 실시간으로 렌더링 하기에는 적합하지 않은 구조이다. 그 이유는 다음과 같다. 먼저 변화하는 BRIR을 계산한 다음 이를 현재 표준의 BRIR 파라미터화부를 사용하여 파라미터화 하는 작업이 실시간으로 수행되어야 한다. 또한 오버랩 없는 프레임(frame) 기반의 처리를 수행하기 때문에 BRIR 필터가 계속 변화할 경우 인접한 프레임간 fade-in, fade-out이 없으면 음질의 열화가 발생할 수 있다.
- [0041] 제안한 구조는 청취자의 움직임 과/혹은 청취자와 오디오 장면(scene)과의 상호작용으로 청취자와의 상대적인 위치가 변화하는 직접음과 반사음에 고속 컨볼루션 기반의 HRTF 컨볼루션 렌더링을 수행하더라도 항상 고정된 위치의 HRTF를 사용하기 때문에 인접한 프레임간의 fade-in, fade-out을 필요로 하지 않는다는 장점을 가진다. 또한 재생/가상 레이아웃을 여러 거리에서 측정한 HRTF 데이터 세트에 DBAP (Distance-based Amplitude Panning) 기법을 사용함으로써 거리 렌더링의 효과를 높일 수 있다.
- [0042] 상기 도 6의 구조 중 HRTF 컨볼루션 연산부는 MPEG-H 3D 오디오 렌더러의 바이노럴 렌더러로 구현될 수 있으며, 이를 제외한 나머지 부분은 도 4의 객체 렌더러 내부에 포함될 수 있다. 객체 렌더러에서는 동일한 음원에 직접음과 반사음에 해당하는 각각의 시간지연과 필터를 적용하고 VBAP (Vector Based Amplitude Panning) 기법 혹은 DBAP 기법을 적용한다. BRIR을 대신 HRTF를 사용한 바이노럴 렌더러를 적용하여 직접음과 반사음을 계산한다. 따라서 제안한 구조는 MPEG-H 3D오디오 표준과 정합되어 MPEG-I phase 1a, 1b에서 필요한 사용자의 3DoF, 3DoF+ 움직임과 오디오 장면과의 상호작용을 반영한 오디오 제공이 가능함.
- [0043] 이상에서는 본 발명을 구체적인 실시예를 통하여 설명하였으나, 당업자라면 본 발명의 취지 및 범위를 벗어나지 않고 수정, 변경을 할 수 있다. 즉, 본 발명은 멀티 오디오 신호에 대한 바이노럴 렌더링의 실시예에 대하여 설명하였지만, 본 발명은 오디오 신호뿐만 아니라 비디오 신호를 포함하는 다양한 멀티미디어 신호에도 동일하게 적용 및 확장 가능하다. 따라서 본 발명의 상세한 설명 및 실시예로부터 본 발명이 속하는 기술분야에 속한 사람이 용이하게 유추할 수 있는 것은 본 발명의 권리범위에 속하는 것으로 해석된다.

도면

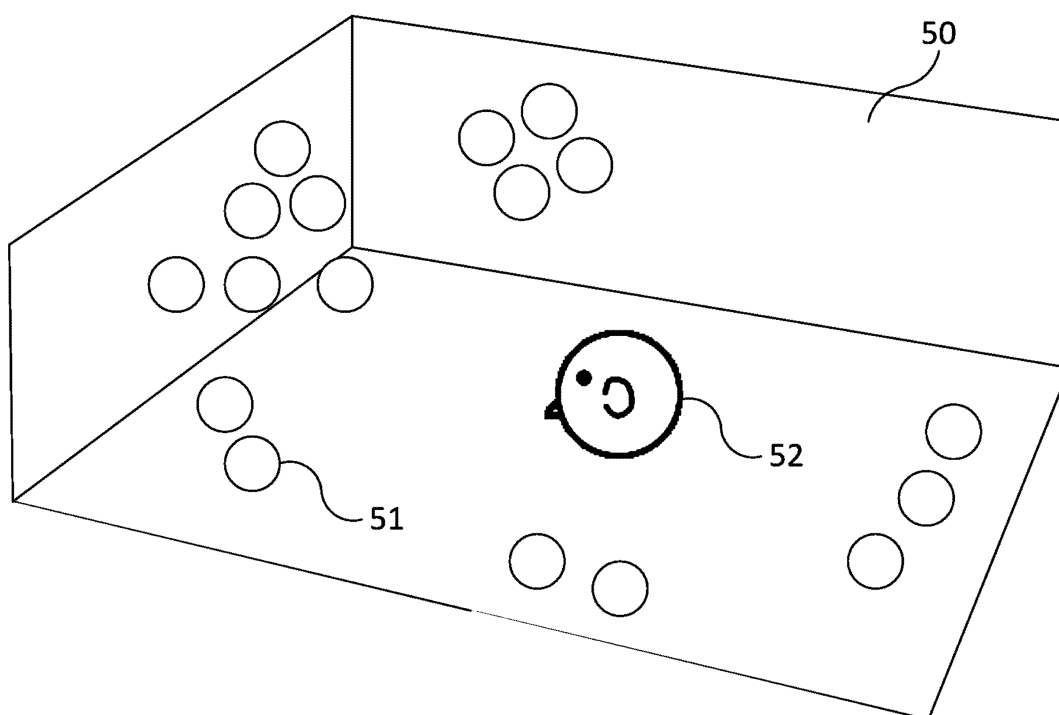
도면1



도면2

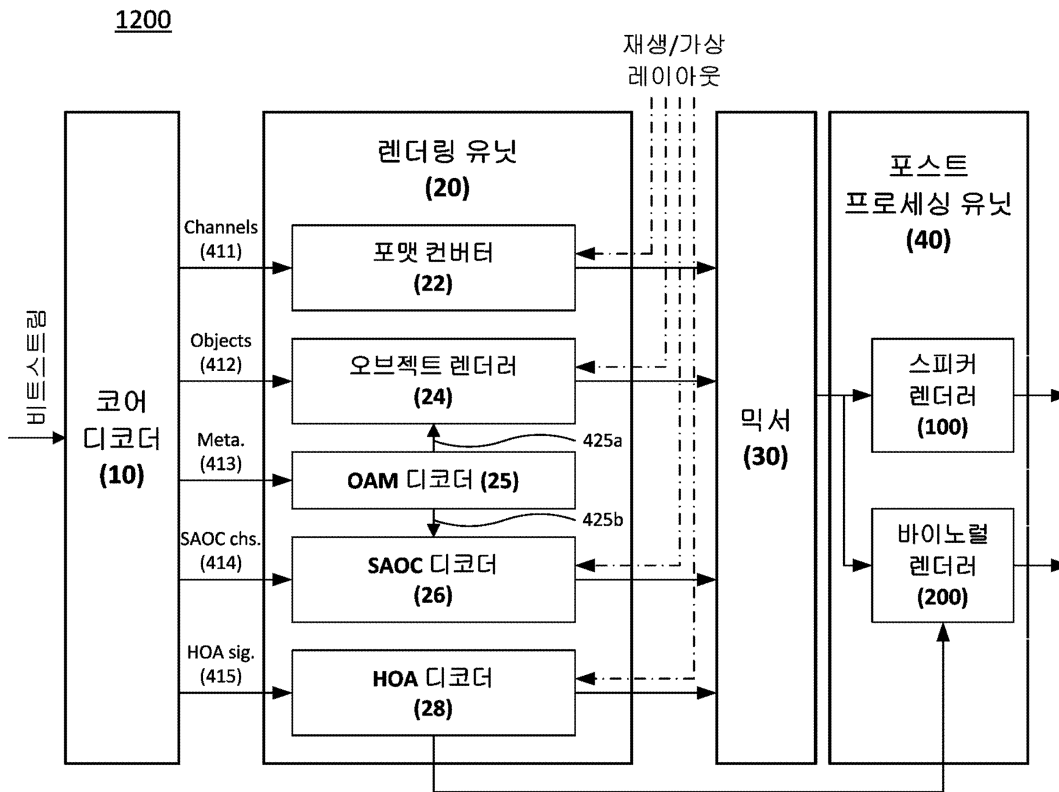


도면3

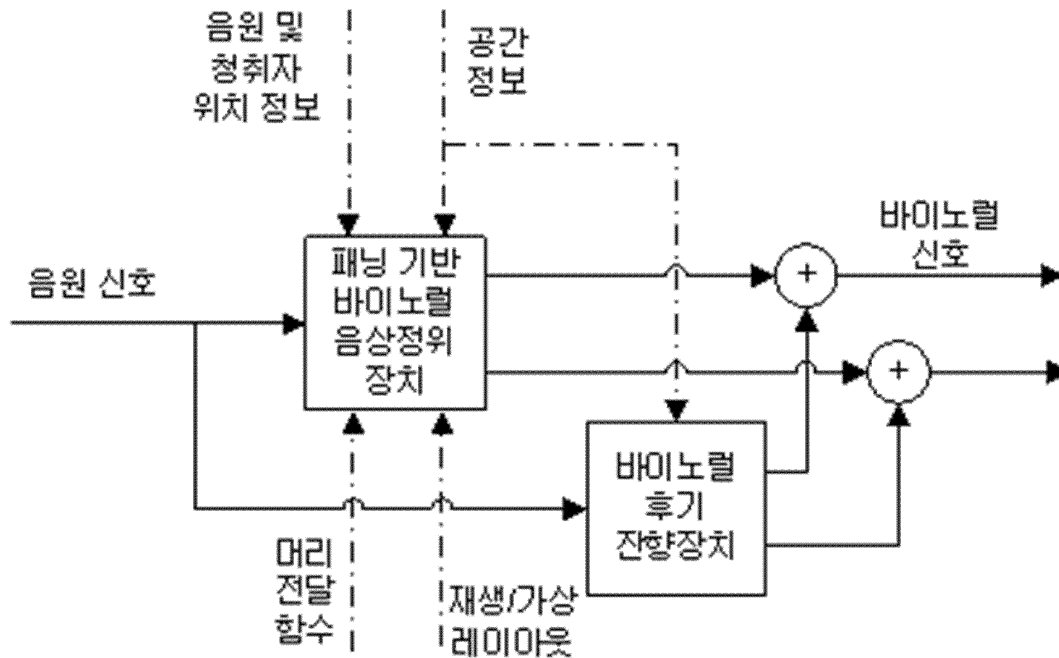




도면4



도면5



도면6

