



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0135604
(43) 공개일자 2017년12월08일

(51) 국제특허분류(Int. Cl.)
H04S 3/00 (2006.01) G10L 19/008 (2014.01)
(52) CPC특허분류
H04S 3/004 (2013.01)
G10L 19/008 (2013.01)
(21) 출원번호 10-2016-0067792
(22) 출원일자 2016년05월31일
심사청구일자 없음

(71) 출원인
가우디오디오랩 주식회사
서울특별시 서초구 마방로 48, 2층(양재동, 정환빌딩)
(72) 발명자
서정훈
서울특별시 강남구 도곡로25길 25, 302호
오현오
경기도 성남시 분당구 동판교로 123 백현마을1단지판교푸르지오그랑블아파트 110-1203
(74) 대리인
임국일

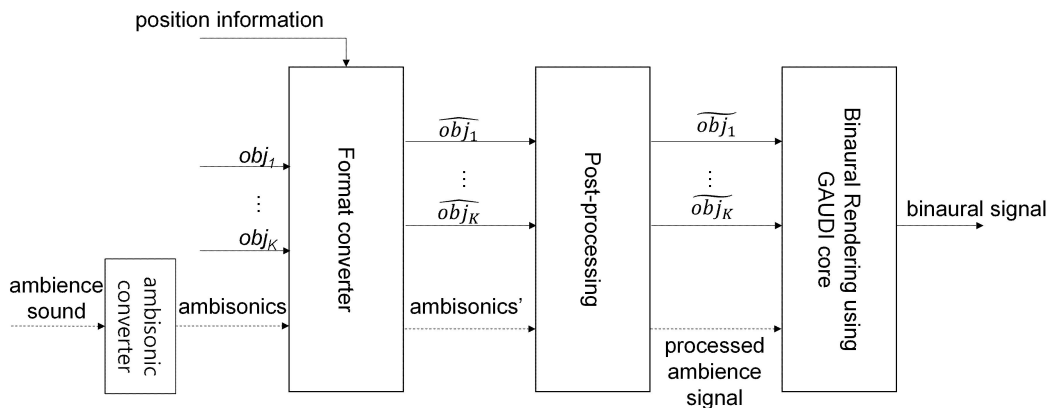
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 오디오 신호 처리 방법 및 장치

(57) 요약

본 발명은 녹음된 신호를 이용하여 DRR matching 된 오디오 신호를 렌더링 하는 방법에 관한 것으로서, 더욱 상세하게는 immersive and interactive audio rendering 을 위한 오디오 신호 처리 방법 및 장치에 관한 것이다.

대표도 - 도1



(52) CPC특허분류

H04S 2400/01 (2013.01)

H04S 2420/01 (2013.01)

(72) 발명자

이태규

서울시 강남구 역삼로20길 10, 403호

전세운

대전광역시 서구 대덕대로 415 상아아파트
106-1405

백용현

서울특별시 서초구 논현로7길 30 일조파크뷰 401

명세서

청구범위

청구항 1

오디오 신호 처리 방법 및 장치.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 신호를 효과적으로 재생하기 위한 신호 처리 방법 및 장치에 관한 것으로서, 더욱 상세하게는 HMD(Head Mounted Display) 기기를 비롯한 휴대 기기를 위한 몰입형(immersive) 바이노럴 렌더링을 구현하기 위한 오디오 신호 처리 방법 및 장치에 관한 것이다.

배경 기술

[0002] HMD에서 immersive audio listening를 하기 위해서 바이노럴 렌더링(binaural rendering) 기술이 필수적이다. 이때, 연산량 및 전력 소모의 제약이 따르는 모바일 디바이스에서, rendering 대상 객체 혹은 채널의 증가로 인한 연산량 및 전력 소모의 부담은 물론이고, 개인 맞춤형으로 제작되지 않은 HRTF 로 인한 개인별 부적응의 문제, HRTF set의 개수 부족 (공간 해상도 부족)에 따른 artifacts, head tracking lag에 따른 성능 저하 및 불편함 등의 문제를 가지고 있다.

발명의 내용

해결하려는 과제

[0003] 본 발명은 상기의 문제점을 해결하기 위해 안출된 것으로, Cinematic 360 VR 에서의 공간감 불일치, 음상정위 왜곡 등의 문제를 해결하고자 하는 목적을 가지고 있다.

과제의 해결 수단

[0004] 본 발명의 실시예에 따르면, 상기와 같은 과제를 해결하기 위한 바이노럴 신호처리 방법 및 장치가 제공될 수 있다.

[0005] << Key Ideas >>

[0006] 1. Cinematic VR의 제작 과정에서 보다 몰입감 있는 경험을 제공하기 위해 오디오 신호를 녹음하는 과정은 매우 중요하다. 개별 sound object만 녹음할 경우 해당 공간의 음향 특성을 반영하기 힘들고, ambience 만 녹음하는 경우 정확한 sound object의 위치를 인지하기 힘든 문제가 있다. 이러한 문제를 해결하기 위해 제작 과정에서 sound object와 ambience를 동시 녹음하여 후처리를 통해 뚜렷한 음상의 위치와 공간감을 동시에 제공할 수 있는 방법이 있다.

[0007] 2. Cinematic 360 VR 에서 효과적인 immersive spatial audio를 재생하기 위해 제작시 개별 sound object와 ambience를 함께 녹음하여 처리할 수 있다. 이 경우 processing 단에서 sound object와 ambience를 함께 재생하는데, object와 ambience의 에너지 비율 차이에 의해 실제 공간의 음향 특성과 다른 mix가 생길 가능성이 있다.

[0008] 3. Sound object를 명확히 녹음하기 위해 사용한 마이크 입력신호는 공간의 잔향을 거의 포함하지 않은 직접음 성분만 포함하고 있는 반면, ambience를 녹음하기 위해 사용한 마이크로폰에는 direct sound와 함께 early reflections, late reverberation 등 공간의 잔향이 함께 포함되어 있으므로 ambience마이크 녹음신호를 이용해 실제 녹음한 공간의 음향 특성을 추출하여 이를 processing 단계에서 사용함으로써 실제 공간의 음향 특성을 보다 사실적으로 반영하여 VR에서 몰입감을 보다 높일 수 있다.

발명의 효과

[0009] 본 발명의 실시예에 따르면, Cinematic 360 VR 제작시 녹음된 sound object 및 앰비언스 신호를 이용하여 보다 immersive한 렌더링을 할 수 있다.

[0010] Sound object는 위치 정보를 이용하여 HoA 및 FoA 신호로 변환될 수 있다.

[0011] 변환된 HoA 및 FoA는 HoA 및 FoA 형태로 변환된 ambience sound와 함께 바이노럴 렌더링되어 immersive binaural rendering을 할 수 있다.

도면의 간단한 설명

[0012] 도 1은 정확한 음상정위와 공간감을 제공하기 위한 cinematic 360 VR오디오의 취득, 처리, 생성과정에 대한 블록도이다.

도 2는 Cinematic 360VR오디오에서 취득 신호에 따른 최종 바이노럴 렌더링의 최종 perceptual evaluation의 결과이다.

도 3은 렌더러 호환성을 위한 HoA+object 와 FoA 변환과정 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0013] 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.

[0015] Cinematic VR의 제작 과정에서 보다 몰입감 있는 경험을 제공하기 위해 오디오 신호를 녹음하는 과정은 매우 중요하다. 개별 sound object만 녹음할 경우 해당 공간의 음향 특성을 반영하기 힘들고, ambience 만 녹음하는 경우 정확한 sound object의 위치를 인지하기 힘든 문제가 있다. 이러한 문제를 해결하기 위해 제작 과정에서 sound object와 ambience를 동시 녹음하여 후처리를 통해 뚜렷한 음상의 위치와 공간감을 동시에 제공할 수 있는 방법이 있다. 해당 과정을 블록도로 나타내면 도 1 과 같다.

[0016] 도 1에서 $obj_1, \dots, obj_K$ 는 녹음시 개별 object로 간주한 음향 신호이고, ambience sound는 개별 사운드 및 녹음 공간에 의해 생성된 앰비언스 마이크로폰 위치에서의 음장 (soundfield) 신호이다. 신호의 흐름을 구별하기 위해 sound object는 실선, ambience에 해당하는 신호는 점선으로 그 흐름을 표시한다. 이렇게 녹음된 신호는 최종 바이노럴 렌더링 방법에 따라 개별 오브젝트와 앰비언스 녹음 신호는 format converter의 과정을 통해 각각 $\overline{obj_1}, \dots, \overline{obj_K}$ 및 앰비소닉 신호로 변환된다. 이 과정에서 $\overline{obj_1}, \dots, \overline{obj_K}$ 는 최종 렌더러에 따라 그 형태가 변할 수 있다. 이 경우 format converter를 거쳐 나오는 신호의 조합, 또는 취득 방법에 따른 신호의 조합은 크게 아래와 같이 분류될 수 있다.

[0017] 1) Sound object + Higher Order Ambisonics (HoA) of Ambience

[0018] 2) 1)의 Sound Object 의 HoA 변환 신호 + HoA of Ambience = Mixed HoA

[0019] 3) 1)의 Sound Object 의 First Order Ambisonics (FoA) + FoA of Ambience = Mixed FoA

[0020] 4) FoA of Ambience

[0022] 이하 설명의 편의를 위해 sound object의 개수가 1개일 때를 가정하여 설명하지만 sound object의 개수가 복수개인 경우에도 동일한 방법에 의해 처리될 수 있다.

[0023] 위의 분류 1)에 마이크로폰 어레이를 이용해 녹음된 신호 pa로부터 구해지는데, 이는 다음 수학적 식 1과 같은 관계를 갖는다.

[0025] (수학식 1)

[0026]

$$p_a(\theta_q, \phi_q) = \sum_{m=0}^{\infty} \sum_{n=-m}^m B^{nm} Y^{nm}(\theta_q, \phi_q)$$

[0028] 수학식 1에서 θ_q 및 ϕ_q 는 마이크론 어레이의 개별 구성 마이크론의 수평각 및 수직각이다. 또한 Y는 해당 수평각, 수직각을 입력으로 하는 구면조화함수 (spherical harmonic functions)이다. m과 n은 구면조화함수의 order와 degree를 의미한다. 위의 수학식 1을 Matrix 형태로 나타내면 수학식 2와 같이 나타낼 수 있고, 수학식 2에서 얻고자 하는 HoA of ambience 신호는 B이다.

[0030] (수학식 2)

[0031]
$$\mathbf{p}_a = \mathbf{YB}$$

[0032]
$$\Leftrightarrow \mathbf{B} = \mathbf{pinv}(\mathbf{Y}) \mathbf{p}_a$$

[0034] 또한 분류 2)에서의 sound object 의 HoA 변환 신호는 기준점에 대한 위치 (θ_s, ϕ_s) 에 대한 구면조화 함수를 이용해 아래 수학식 3과 같이 표현할 수 있다.

[0036] (수학식 3)

[0037]
$$B_{nm}^s = SY(\theta_s, \phi_s)$$

[0039] 해당 사운드 오브젝트의 위치는 개별 마이크론과 기준점에 설치된 외부 센서를 이용하여 측정하거나 마이크론 어레이의 신호를 분석하여 추정한다.

[0040] 무한대의 구면조화 함수 차수를 이용할 수 없는 현실적인 조건들을 고려하여 앰비소닉 신호는 구면조화함수의 order를 M차수로 truncation하여 근사화한다. 또한 수학식 2와 수학식 3을 이용해 사운드 오브젝트와 ambience에 대한 HoA 신호를 얻었다면 0 차수와 1차수 성분은 남기고 고차 성분들을 삭제하여 FoA신호를 얻을 수 있다. 이 근사화 과정에서 공간 해상도 (spatial resolution) 역시 저하되고, M 이 작을수록 이 열화는 심해진다. 이러한 이유로 binaural rendering 시의 성능이 사운드 오브젝트는 object-based rendering 을 하고, 앰비언스 신호는 scene-based rendering을 하는 것이 가장 좋은 성능을 보여준다. 분류 1)에 기반한 binaural rendering의 성능을 기준으로 했을 때 분류 2), 3) 및 sound object는 제외하고 scene-based 렌더링을 한 경우의 인지적 주관평가 결과는 도 2와 같다.

[0042] 도 2에서 확인할 수 있듯이 분류 1)을 기준으로 하였을 때 최종 렌더링 품질은 분류 2) > 분류 3) 으로 나타낼 수 있고 FoA 앰비언스 신호만 scene-based 렌더링을 했을 때의 품질이 가장 낮음을 알 수 있다. FoA 기반의 렌더러의 대표적인 예로 현재 YouTube Spatial Audio를 들 수 있다. 분류 1) 신호와 분류 3) 또는 4)에 해당하는 신호는 렌더러 호환성 (e.g. YouTube Spatial Audio) 을 위한 format conversion 이 가능한데, 이는 도 3과 같이 나타낼 수 있다.

[0044] 도 3에서 Binaural_1 신호는 신호분류 1)을 이용하여 GAUDIO Renderer에 의해 바이노럴 렌더링 된 신호, Binaural_2 신호는 신호분류 3) 또는 4) 를 이용하여 format conversion 과정을 거친 후 GAUDIO Renderer에 의해 바이노럴 렌더링 된 신호, Binaural_3은 신호분류 1)을 이용하였지만 format conversion을 거친 후 other

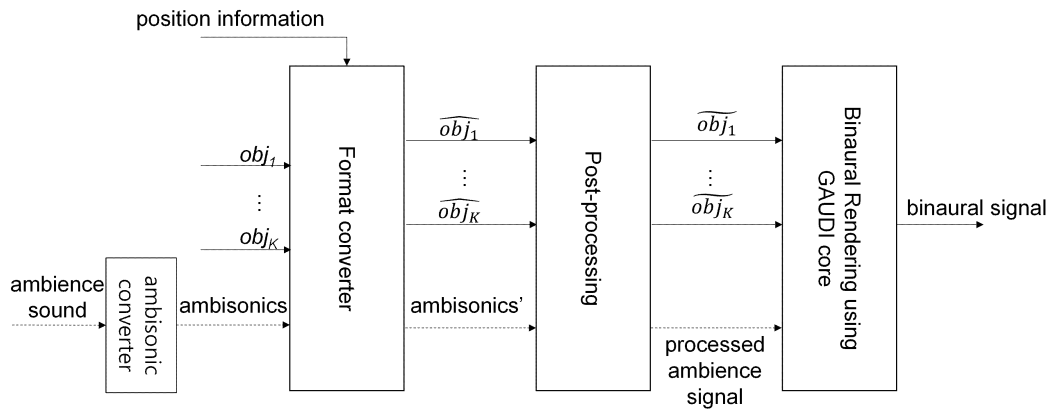
renderer에 의해 바이노럴 렌더링 된 신호, 그리고 Binaural_4는 신호분류 3) 또는 4) 를 이용하여 other renderer에 의해 바이노럴 렌더링 된 신호를 나타낸다. 또한 실선 및 점선은 신호의 흐름을 볼 때 최초 생성 음원이 무엇인지를 나타내는데, 실선은 HoA+Object 신호로부터, 점선은 FoA 신호로부터 생성된 신호의 흐름을 나타낸다.

- [0046] [what is claimed here]:
- [0047] 1.
- [0048] hoa(foa)에 해당하는 제1 오디오 신호를 수신
- [0049] object에 해당하는 제2오디오 신호를 수신
- [0050] hoa 가 구성한 sound scene 상에서 상기 object 의 위치 정보를 수신 (직접 계산하거나 외부의 센서를 이용하여 identify하여 수신)
- [0051] 상기 object 위치 정보를 이용하여, 제1신호와 제2신호를 합성한 제3신호를 생성하는 방법
- [0052] 2.
- [0053] 상기 제3신호는 hoa 신호인 것을 특징
- [0054] 3.
- [0055] 상기 제3신호는 다시 foa로 변환되어 재생(렌더링)되는 것을 특징
- [0056] 4.
- [0057] 제1신호, 제2신호 및 위치 정보를 비트열로 생성하여 전송하고, 상기 제3신호를 생성하는 과정은 수신단에서 이뤄지는 것을 특징
- [0058] 5.
- [0059] 상기 제3신호를 생성하는 방법에 있어서, 상기 제1신호를 foa로 변환, 제2신호를 foa로 변환하여 변환된 두 foa를 더하는 방법
- [0060] 6.
- [0061] 상기 제3신호를 생성함에 있어서, 상기 제1신호와 제2신호를 먼저 hoa에서 합성하고, 이를 foa로 변환하는 방법

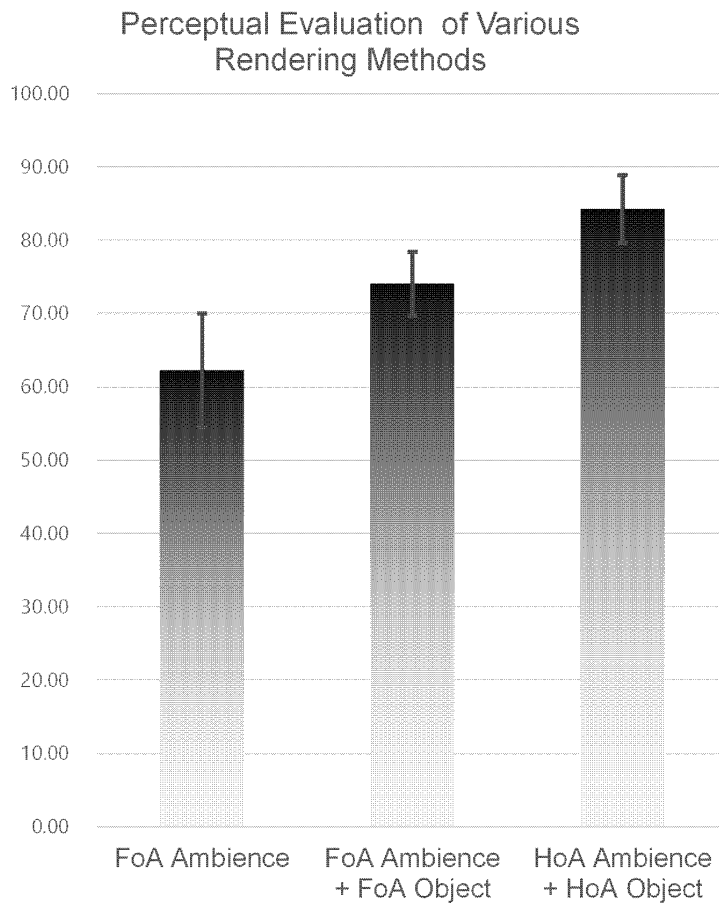
부호의 설명

도면

도면1



도면2



도면3

