



등록특허 10-2187936



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년12월07일
(11) 등록번호 10-2187936
(24) 등록일자 2020년12월01일

- (51) 국제특허분류(Int. Cl.)
G10L 19/008 (2014.01) *H04S 3/02* (2006.01)
- (52) CPC특허분류
G10L 19/008 (2020.08)
H04S 3/02 (2013.01)
- (21) 출원번호 10-2020-7017672(분할)
- (22) 출원일자(국제) 2013년07월16일
심사청구일자 2020년07월17일
- (85) 번역문제출일자 2020년06월18일
- (65) 공개번호 10-2020-0077601
- (43) 공개일자 2020년06월30일
- (62) 원출원 특허 10-2015-7000876
원출원일자(국제) 2013년07월16일
심사청구일자 2018년07월09일
- (86) 국제출원번호 PCT/EP2013/065032
- (87) 국제공개번호 WO 2014/012944
국제공개일자 2014년01월23일

(30) 우선권주장
12305861.2 2012년07월16일
유럽특허청(EPO)(EP)

(56) 선행기술조사문현

EP02469741 A1

(뒷면에 계속)

전체 청구항 수 : 총 6 항

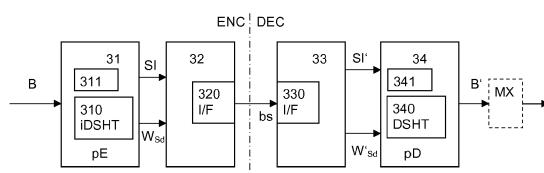
심사관 : 이남숙

(54) 발명의 명칭 잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 방법 및 장치와, 잡음 감소를 위한
다채널 HOA 오디오 신호를 디코딩하는 방법 및 장치

(57) 요 약

잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 방법은 역 적응 DSHT를 이용하여 채널을 역상관(decorrelate)하는 단계(81) - 상기 역 적응 DSHT는 회전 동작(330) 및 역 DSHT(810)를 포함하고, 상기 회전 동작은 상기 iDSHT의 공간 샘플링 그리드를 회전함 - , 상기 역상관된 채널의 각각을 지각적으로 인코딩하는 단계(82), 회전 정보(SI)를 인코딩하는 단계 - 상기 회전 정보는 상기 회전 동작을 정의하는 파라미터를 포함함 - , 및 상기 지각적으로 인코딩된 오디오 채널 및 상기 인코딩된 회전 정보를 송신 또는 저장하는 단계를 포함한다.

대 표 도 - 도3



(72) 발명자

크뢰거, 알렉산더

독일 30625 하노버 칼 비헤르트 알레 74 도이치 톰
슨 오에이치쥐 리서치 앤드 이노베이션

작스, 피터

독일 30625 하노버 칼 비헤르트 알레 74 도이치 톰
슨 오에이치쥐 리서치 앤드 이노베이션

(56) 선행기술조사문헌

EP02469742 A2

KR1020120070521 A

Jorge TREVINO, et al. High order Ambisonic decoding method for irregular loudspeaker arrays. Proceedings of 20th International Congress on Acoustics. 2010. pp. 23-27.

Robert E. Davis, et al. A Simple and Efficient Method for Real-Time Computation and Transformation of Spherical Harmonic based Sound Fields. Audio Engineering Society. 2012.10.26.

명세서

청구범위

청구항 1

인코딩된 HOA(higher order ambisonics) 오디오 신호들을 디코딩하는 방법으로서,

상기 인코딩된 HOA 오디오 신호들 및 회전 정보를 수신하는 단계;

상기 인코딩된 HOA 오디오 신호들에 대응하는 HOA 표현들을 결정하기 위해 지각적인 디코딩에 기초하여 상기 인코딩된 HOA 오디오 신호들을 압축 해제하는 단계;

상기 회전 정보와 연관된 구면 샘플 그리드의 회전에 기초하여 회전된 변환을 결정하는 단계; 및

상기 회전된 변환 및 상기 HOA 표현에 기초하여 회전된 HOA 표현을 결정하는 단계

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 회전된 변환은:

디폴트 구면 샘플 그리드를 선택하고;

M개의 시간 샘플들의 블록에 대하여, 회전된 구면 샘플 그리드를 결정하기 위해 회전 정보에 기초하여 상기 디폴트 구면 샘플 그리드를 회전시키고; 그리고

상기 회전된 구면 샘플 그리드에 대한 모드 매트릭스를 결정하는 것에 기초하여 결정되는, 방법.

청구항 3

제1항에 있어서,

상기 회전 정보는 3개의 각($\theta_{axis}, \phi_{axis}, \varphi_{rot}$)에 기초한 3개의 성분 회전에 대응하고, 여기서,

$\theta_{axis}, \phi_{axis}$ 는 구면 좌표 내의 1의 암시적 반경을 갖는 회전 축에 대한 정보를 정의하고, φ_{rot} 는 상기 회전 축 주변의 회전 각을 정의하는, 방법.

청구항 4

인코딩된 HOA 오디오 신호들을 디코딩하는 장치로서,

상기 인코딩된 HOA 오디오 신호들 및 회전 정보를 수신하기 위한 수신기; 및

디코더를 포함하고, 상기 디코더는:

상기 인코딩된 HOA 오디오 신호들에 대응하는 HOA 표현들을 결정하기 위해 지각적인 디코딩에 기초하여 상기 인코딩된 HOA 오디오 신호들을 압축 해제하고;

상기 회전 정보와 연관된 구면 샘플 그리드의 회전에 기초하여 회전된 변환을 결정하고; 그리고

상기 회전된 변환 및 상기 HOA 표현에 기초하여 회전된 HOA 표현을 결정하도록 구성되는, 장치.

청구항 5

제4항에 있어서,

상기 디코더는 새로운 변환을 위한 디폴트 구면 샘플 그리드의 선택; M개의 시간 샘플들의 블록에 대하여, 회전된 구면 샘플 그리드를 결정하기 위한 상기 회전 정보에 따른 상기 디폴트 구면 샘플 그리드의 회전; 및 상기

회전된 구면 샘플 그리드에 대한 모드 매트릭스의 결정에 기초하여 상기 회전된 변환을 결정하도록 구성되는, 장치.

청구항 6

제4항에 있어서, 상기 회전 정보는 3개의 각($\theta_{axis}, \phi_{axis}, \varphi_{rot}$)에 기초한 3개의 컴포넌트 회전에 대응하고, 여기서, $\theta_{axis}, \phi_{axis}$ 는 구면 좌표 내의 1의 암시적 반경을 갖는 회전 축에 대한 정보를 정의하고, φ_{rot} 는 상기 회전 축 주변의 회전 각을 정의하는, 장치.

발명의 설명

기술 분야

[0001] 본 발명은 잡음 감소를 위한 다채널 HOA(higher order ambisonics) 오디오 신호를 인코딩하는 방법 및 장치와, 잡음 감소를 위한 다채널 HOA 오디오 신호를 디코딩하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] HOA는 다채널 음장 표현[4]이고 HOA 신호는 다채널 오디오 신호이다. 특정한 확성기 셋업 상의 소정의 다채널 오디오 신호 표현, 특히 HOA 표현의 재생은 통상 매트릭스화 동작으로 구성된 특수 렌더링을 요구한다. 디코딩 후, 앰비소닉스(Ambisonics) 신호는 "매트릭스화"되고, 즉, 예를 들어 확성기의 실제 공간 위치에 대응하는 새로운 오디오 신호에 맵핑된다. 통상, 단일 채널들 간의 높은 상호 상관이 존재한다.

[0003] 매트릭스화 동작후 코딩 잡음의 증가를 경험한다는 문제가 있다. 그 이유가 종래 기술에는 잘 알려지지 않은 것으로 나타난다. 이 효과는 또한 지각적 코더(perceptual coder)에 의한 압축 전에, 예를 들어, DSHT(discrete spherical harmonics transform)에 의해 HOA 신호를 공간 도메인으로 변형할 때 발생한다.

[0004] HOA 오디오 신호 표현의 압축을 위한 통상의 방법은 개별 앰비소닉스 계수 채널에 독립적인 지각적 코더를 적용하는 것이다[7]. 특히, 지각적 코더는 단지 각각의 개별 단일 채널 신호 내에서 발생하는 코딩 잡음 마스킹 효과를 고려한다. 그러나, 이러한 효과는 일반적으로 비선형이다. 이러한 단일 채널을 새로운 신호로 매트릭스화하면, 잡음 언마스킹(noise unmasking)이 발생할 가능성이 있다. 이 효과는 또한 HOA 신호가 지각적 코더에 의한 압축 전에 DSHT에 의해 공간 도메인으로 변형될 때 발생한다[8].

[0005] 이러한 다채널 오디오 신호 표현의 송신 또는 저장은 통상 적절한 다채널 압축 기술을 요구한다. 통상, 채널 독립 지각적 디코딩은 I개의 디코딩 신호($\hat{\mathbf{x}}_i(l), i = 1, \dots, I$)를 J개의 새로운 신호($\hat{\mathbf{y}}_j(l), j = 1, \dots, J$)로 최종적으로 매트릭스화하기 전에 수행된다. 매트릭스화라는 용어는 디코딩 신호($\hat{\mathbf{x}}_i(l)$)를 가중 방식으로 가산 또는 혼합하는 것을 의미한다.

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

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

[0006] [0007] 예 따라 모든 신호($\hat{\mathbf{x}}_i(l), i = 1, \dots, I$) 뿐만 아니라 모든 새로운 신호($\hat{\mathbf{y}}_j(l), j = 1, \dots, J$)를 벡터에 배치하면, "매트릭스화"라는 용어는 $\hat{\mathbf{y}}(l)$ 이 매트릭스 동작

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

[0009] 을 통해 $\hat{\mathbf{x}}(l)$ 로부터 수학적으로 얻어진다는 사실로부터 기인하고, 여기서 A는 혼합 가중치들로 구성된 혼합 매트릭스를 나타낸다. "혼합" 및 "매트릭스화"라는 용어는 여기에 동의어로 사용된다. 혼합/매트릭스화는 임의의 특정 확성기 셋업을 위한 오디오 신호를 렌더링할 목적으로 사용된다. 매트릭스가 의존하는 특정 개별 확성기 셋업 및 렌더링 동안 매트릭스화를 위해 사용되는 매트릭스는 통상 지각적인 코딩 스테이지에서 알려져 있지 않다.

발명의 내용

해결하려는 과제

[0010] 본 발명은 다채널 HOA 오디오 신호를 인코딩 및/또는 디코딩하여 잡음 감소를 얻는 개선책을 제공한다. 특히, 본 발명은 3D 오디오 레이트 압축을 위한 코딩 잡음 디마스킹(coding noise demasking)을 억제하는 방법을 제공한다.

[0011] 본 발명은 (원치 않는) 잡음 언마스킹 효과를 최소화하는 aDSHT(adaptive Discrete Spherical Harmonics Transform)에 대한 기술을 기재한다. 또한, aDSHT가 압축 코더 아키텍처 내에서 통합될 수 있는 방법을 기재한다. 기재된 기술은 특히 적어도 HOA 신호에 대하여 유리하다. 본 발명의 하나의 이점은 송신될 사이드 정보의 양이 감소되는 것이다. 원리적으로, 회전 축 및 회전 각만이 송신될 필요가 있다. DSHT 샘플링 그리드는 송신되는 채널의 수만큼 간접적으로 시그널링될 수 있다. 이 사이드 정보의 양은 상관 매트릭스의 절반보다 많이 송신될 필요가 있는 KLT(Karhunen Loeve transform) 같은 다른 어프로치와 비교하여 매우 작다.

과제의 해결 수단

[0012] 본 발명의 일 실시예에 따르면, 잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 방법은 역 적응 DSHT를 이용하여 채널을 역상관(decorrelate)하는 단계 - 상기 역 적응 DSHT는 회전 동작 및 역 DSHT(iDSHT)를 포함하고, 상기 회전 동작은 상기 iDSHT의 공간 샘플링 그리드를 회전함 -, 상기 역상관된 채널의 각각을 지각적으로 인코딩하는 단계, 회전 정보를 인코딩하는 단계 - 상기 회전 정보는 상기 회전 동작을 정의하는 파라미터를 포함함 -, 및 상기 지각적으로 인코딩된 오디오 채널 및 상기 인코딩된 회전 정보를 송신 또는 저장하는 단계를 포함한다. 역 적응 DSHT를 이용하여 채널을 역상관하는 단계는 원리적으로 공간 인코딩 단계이다.

[0013] 본 발명의 일 실시예에 따르면, 감소된 잡음을 갖는 코딩된 다채널 HOA 오디오 신호를 디코딩하는 방법은 인코딩된 다채널 HOA 오디오 신호 및 채널 회전 정보를 수신하는 단계, 상기 수신된 데이터를 압축 해제하는 단계 - 지각적인 디코딩이 사용됨 -, 적응 DSHT(aDSHT)를 이용하여 각각의 채널을 공간적으로 디코딩하는 단계, 지각적으로 및 공간적으로 디코딩된 채널을 상관하는 단계 - 상기 회전 정보에 따른 상기 aDSHT의 공간 샘플링 그리드의 회전이 수행됨 -, 및 상기 지각적으로 및 공간적으로 디코딩되고 상관된 채널을 매트릭스화하는 단계 - 확성기 위치에 맵핑된 재생가능한 오디오 신호가 얻어짐 - 를 포함한다.

[0014] 다채널 HOA 오디오 신호를 인코딩하는 장치가 청구항 11에 개시된다. 다채널 HOA 오디오 신호를 디코딩하는 장치가 청구항 12에 개시된다.

[0015] 일 형태에 있어서, 컴퓨터 판독가능 매체는 컴퓨터가 상술한 단계를 포함하는 인코딩 방법 또는 상술한 단계를 포함하는 디코딩 방법을 수행하도록 하는 실행가능한 명령을 갖는다. 본 발명의 유리한 실시예는 종속 청구항, 다음의 설명 및 도면에 개시된다.

도면의 간단한 설명

[0016] 본 발명의 예시적인 실시예는 첨부된 도면을 참조하여 설명한다.

도 1은 M개의 계수의 블록을 레이트 압축하는 기지의 인코더 및 디코더를 나타내는 도면.

도 2는 종래의 DSHT(discrete spherical harmonics transform) 및 종래의 역 DSHT를 이용하여 HOA 신호를 공간 도메인으로 변환하는 기지의 인코더 및 디코더를 나타내는 도면.

도 3은 적응 DSHT 및 적응 역 DSHT를 이용하여 HOA 신호를 공간 도메인으로 변환하는 인코더 및 디코더를 나타내는 도면.

도 4는 테스트 신호를 나타내는 도면.

도 5는 인코더 및 디코더 형성 블록에 사용되는 코드북을 위한 구면 샘플링 위치의 예를 나타내는 도면.

도 6은 신호 적응 DSHT 형성 블록(pE 및 pD)를 나타내는 도면.

도 7은 본 발명의 제1 실시예를 나타내는 도면.

도 8은 인코딩 프로세스 및 디코딩 프로세스의 플로우챠트.

도 9는 본 발명의 제2 실시예를 나타내는 도면.

발명을 실시하기 위한 구체적인 내용

[0017]

도 2는 역 DSHT를 이용하여 HOA 신호가 공간 도메인으로 변환되는 기지의 시스템을 나타낸다. 신호는 iDSHT(21), 레이트 압축(E1)/압축해제(D1)를 이용하여 변환되고 DSHT(24)를 이용하여 계수 도메인(S24)으로 재변환된다. 이와 다르게, 도 3은 본 발명의 일 실시예에 따른 시스템을 나타낸다. 기지의 솔루션의 DSHT 처리 블록은 역 적응 DSHT 및 적응 DSHT를 각각 제어하는 처리 블록(31, 34)으로 대체된다. 사이드 정보(SI)는 비트 스트림(bs) 내에서 송신된다. 시스템은 다채널 HOA 오디오 신호를 인코딩하는 장치의 엘리먼트 및 다채널 HOA 오디오 신호를 디코딩하는 장치의 엘리먼트를 포함한다.

[0018]

일 실시예에서, 잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 장치(ENC)는 역 적응 DSHT(i aDSHT)를 이용하여 채널(B)을 역상관(decorrelate)하는 역상관기(31)를 포함하고, 역 적응 DSHT는 회전 동작 유닛(311) 및 역 DSHT(iDSHT)(310)를 포함한다. 회전 동작 유닛은 iDSHT의 공간 샘플링 그리드를 회전한다. 역상관기(31)는 역상관된 채널(W_{sd}) 및 회전 정보를 포함하는 사이드 정보(SI)를 제공한다. 또한, 장치는 역상관된 채널(W_{sd})의 각각을 지각적으로 인코딩하는 지각적 인코더(32) 및 회전 정보를 인코딩하는 사이드 정보 인코더(321)를 포함한다. 회전 정보는 상기 회전 동작을 정의하는 파라미터를 포함한다. 지각적 인코더(32)는 지각적으로 인코딩된 오디오 채널 및 인코딩된 회전 정보를 제공하여 데이터 레이트를 감소시킨다. 마지막으로, 인코딩 장치는, 지각적으로 인코딩된 오디오 채널 및 인코딩된 회전 정보로부터 비트스트림(bs)을 생성하고 비스스트림(bs)를 송신 또는 저장하는 인터페이스 수단(320)을 포함한다.

[0019]

감소된 잡음을 갖는 다채널 HOA 오디오 신호를 디코딩하는 장치(DEC)는 인코딩된 다채널 HOA 오디오 신호 및 채널 회전 정보를 수신하는 인터페이스 수단(330) 및 수신된 데이터를 압축 해제하고 각각의 채널을 지각적으로 디코딩하는 지각적 디코더를 포함하는 압축 해제 모듈(33)을 포함한다. 압축 해제 모듈(33)은 지각적으로 디코딩되고 회복된 채널(W'_{sd}) 및 회복된 사이드 정보(SI')를 제공한다. 또한, 디코딩 장치는 적응 DSHT(aDSHT)를 이용하여 지각적으로 디코딩된 채널(W'_{sd})을 상관하는 상관기(34) - DSHT 및 상기 회전 정보에 따른 DSHT의 공간 샘플링 그리드의 회전이 수행됨 - 및 지각적으로 디코딩되고 상관된 채널을 매트릭스화하는 믹서(MX)- 확성기 위치에 맵핑된 재생가능 오디오 신호가 얻어짐 - 를 포함한다. 적어도 aDSHT는 상관기(34) 내의 DSHT 유닛(340)에서 수행될 수 있다. 일 실시예에서, 공간 샘플링 그리드의 회전은, 원리적으로 본래의 DSHT 샘플링 포인트를 재산출하는 그리드 회전 유닛(341)에서 수행된다. 다른 실시예에서, 회전은 DSHT 유닛(340) 내에서 수행된다.

[0020]

다음에서, 언마스킹(unmasking)을 정의하고 기술하는 수학적 모델이 주어진다. 주어진 이산 시간 다채널 신호가 I개의 채널($x_i(m)$, $i = 1, \dots, I$)로 구성되는 것으로 가정하고, 여기서, m 은 시간 샘플 인덱스를 나타낸다. 개별 신호는 실수(real value) 또는 복소수일 수 있다. 시간 샘플 인덱스($m_{\text{START}} + 1$)에서 시작하는 M개의 샘플의 프레임을 고려하고, 개별 신호는 정지된 것으로 가정한다. 해당 샘플은

수학식 1

$$\mathbf{X} := [\mathbf{x}(m_{\text{START}} + 1), \dots, \mathbf{x}(m_{\text{START}} + M)]$$

[0021]

에 따라 매트릭스($\mathbf{X} \in \mathbb{C}^{I \times M}$) 내에 배치되고,

[0023]

여기서,

수학식 2

$$\mathbf{x}(l) := [x_1(m), \dots, x_l(m)]^T$$

[0024]

$(\cdot)^T$ 는 전치(transposition)를 나타낸다. 해당 경험 상관 매트릭스(empirical correlation matrix)는

수학식 3

$$\boldsymbol{\Sigma}_{\mathbf{X}} := \mathbf{X} \mathbf{X}^H$$

[0026]

으로 주어지고, 여기서,

$(\cdot)^H$ 는 조인트 복소 공액 및 전치(joint complex conjugation and transposition)를 나타낸다.

[0029]

지금부터, 다채널 신호 프레임이 코딩된다고 가정하면, 이로써 복원시에 코딩 에러 잡음이 유입된다. 따라서,

$\widehat{\mathbf{X}}$ 로 표시된 복원된 프레임 샘플의 매트릭스는

수학식 4

$$\widehat{\mathbf{X}} = \mathbf{X} + \mathbf{E}$$

[0030]

에 따라 정확한(true) 샘플 매트릭스(X) 및 코딩 잡음 컴포넌트(E)로 구성되고,

[0032]

여기서,

수학식 5

$$\mathbf{E} := [\mathbf{e}(m_{\text{START}} + 1), \dots, \mathbf{e}(m_{\text{START}} + L)]$$

[0033]

및

수학식 6

$$\mathbf{e}(m) := [e_1(m), \dots, e_l(m)]^T$$

[0035]

이다.

[0036]

각각의 채널이 독립적으로 코딩된 것으로 가정되므로, 코딩 잡음 신호($e_i(m)$)는 $i=1, \dots, l$ 에 대하여 서로 독립적인 것으로 가정할 수 있다. 잡음 신호는 제로 평균이라는 특성 및 가정을 이용하여, 잡음 신호의 경험 상관 매트릭스는 다음과 같이 대각선 매트릭스로 주어진다.

수학식 7

$$\Sigma_E = \text{diag}(\sigma_{e_1}^2, \dots, \sigma_{e_I}^2)$$

[0038] 여기서, $\text{diag}(\sigma_{e_1}^2, \dots, \sigma_{e_I}^2)$ 는 대각선 상의 경험 잡음 신호 전력

수학식 8

$$\sigma_{e_i}^2 = \frac{1}{M} \sum_{m=m_{\text{START}}+1}^{m_{\text{START}}+M} |e_i(m)|^2$$

[0040] [0041] 을 갖는 대각 매트릭스를 나타낸다. 추가의 필수적인 가정은, 미리 정의된 신호 대 잡음비(SNR)가 각 채널에 대하여 만족하도록 코딩이 수행된다는 것이다. 일반성을 잃지 않고, 미리 정의된 SNR이 각각의 채널에 대하여 동일한 것으로 가정, 즉, 모든 $i = 1, \dots, I$ 에 대하여,

수학식 9

$$\text{SNR}_x = \frac{\sigma_{x_i}^2}{\sigma_{e_i}^2}$$

[0042] [0043] 이고,

수학식 10

$$\sigma_{x_i}^2 := \frac{1}{M} \sum_{m=m_{\text{START}}+1}^{m_{\text{START}}+M} |x_i(m)|^2$$

[0044] [0045] 이다.

[0046] 지금부터, 복원된 신호를 J개의 새로운 신호($y_j(m), j = 1, \dots, J$)로 매트릭스화하는 것을 고려한다. 임의의 코딩 에러를 유입하지 않고, 매트릭스화된 신호의 샘플 매트릭스는

수학식 11

$$\mathbf{Y} = \mathbf{A} \mathbf{X}$$

[0047] [0048] 로 표현될 수 있고,

[0049] 여기서, $\mathbf{A} \in \mathbb{C}^{J \times I}$ 는 혼합 매트릭스를 나타내고, 여기서,

수학식 12

$$[0050] \quad \mathbf{Y} := [\mathbf{y}(m_{\text{START}} + 1), \dots, \mathbf{y}(m_{\text{START}} + M)]$$

[0051] 이고,

수학식 13

$$[0052] \quad \mathbf{y}(m) := [y_1(m), \dots, y_J(m)]^T$$

[0053] 이다.

[0054] 그러나, 코딩 잡음 때문에, 매트릭스화된 신호의 샘플 매트릭스는

수학식 14

$$[0055] \quad \widehat{\mathbf{Y}} := \mathbf{Y} + \mathbf{N}$$

[0056] 로 주어지고, 여기서, \mathbf{N} 은 매트릭스화된 잡음 신호의 샘플을 포함하는 매트릭스이다. 이는

수학식 15

$$[0057] \quad \mathbf{N} = \mathbf{A}\mathbf{E}$$

수학식 16

$$[0058] \quad \mathbf{N} = [\mathbf{n}(m_{\text{START}} + 1) \dots \mathbf{n}(m_{\text{START}} + M)]$$

[0059] 와 같이 표현될 수 있고,

[0060] 여기서,

수학식 17

$$[0061] \quad \mathbf{n}(m) := [n_1(m) \dots n_J(m)]^T$$

[0062] 는 시간 샘플 인덱스(m)에서의 모든 매트릭스화된 잡음 신호의 벡터이다.

[0063] 수학식 11을 이용하여, 매트릭스화된 무잡음 신호의 경험 상관 매트릭스는 다음과 같이 표현될 수 있다.

수학식 18

$$\Sigma_Y = \mathbf{A} \Sigma_X \mathbf{A}^H$$

[0064] 따라서, Σ_Y 의 대각선 상의 j번째 엘리먼트인 j번째 매트릭스화된 무잡음 신호의 경험 전력은 다음과 같이 기재되고,

수학식 19

$$\sigma_{y_j}^2 = \mathbf{a}_j^H \Sigma_X \mathbf{a}_j$$

[0066] 여기서, \mathbf{a}_j 는

수학식 20

$$\mathbf{A}^H = [\mathbf{a}_1, \dots, \mathbf{a}_J]$$

[0067] 예 따른 \mathbf{A}^H 의 j번째 열이다.

[0068] 마찬가지로, 수학식 15로, 매트릭스화된 잡음 신호의 경험 상관 매트릭스는 다음과 같이 기재된다.

수학식 21

$$\Sigma_N = \mathbf{A} \Sigma_E \mathbf{A}^H$$

[0069] Σ_N 의 대각선 상의 j번째 엘리먼트인 j번째 매트릭스화된 잡음 신호의 경험 전력은 다음과 같이 주어진다.

수학식 22

$$\sigma_{n_j}^2 = \mathbf{a}_j^H \Sigma_E \mathbf{a}_j$$

[0070] 결과적으로,

수학식 23

$$\text{SNR}_{y_j} := \frac{\sigma_{y_j}^2}{\sigma_{n_j}^2}$$

[0076] 예 의해 정의된 매트릭스화된 신호의 경험 SNR은

수학식 24

$$\text{SNR}_{y_j} = \frac{\mathbf{a}_j^H \Sigma_{\mathbf{X}} \mathbf{a}_j}{\mathbf{a}_j^H \Sigma_{\mathbf{E}} \mathbf{a}_j}$$

[0077]

[0078]로서 수학식 19 및 수학식 22를 이용하여 다시 표현될 수 있다.

수학식 25

$$\Sigma_{\mathbf{X}} = \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_I}^2) + \Sigma_{\mathbf{X}, \text{NG}}$$

[0079]

[0080]여기서,

수학식 26

$$\Sigma_{\mathbf{X}, \text{NG}} := \Sigma_{\mathbf{X}} - \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_I}^2)$$

[0081]

[0082]로서 $\Sigma_{\mathbf{X}}$ 를 대각선 및 비대각선 성분으로 분해하고,

[0083]수학식 7 및 수학식 9가 모든 채널에 대해 상수 $\text{SNR}(\text{SNR}_x)$ 을 갖는다는 가정으로부터 기인하는 특성

수학식 27

$$\text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_I}^2) = \text{SNR}_x \cdot \text{diag}(\sigma_{e_1}^2, \dots, \sigma_{e_I}^2)$$

[0084]

[0085]을 이용함으로써, 결국 매트릭스화된 신호의 경험 SNR에 대한 원하는 수학식을 얻는다:

수학식 28

$$\text{SNR}_{y_j} = \frac{\mathbf{a}_j^H \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_I}^2) \mathbf{a}_j}{\mathbf{a}_j^H \Sigma_{\mathbf{E}} \mathbf{a}_j} + \frac{\mathbf{a}_j^H \Sigma_{\mathbf{X}, \text{NG}} \mathbf{a}_j}{\mathbf{a}_j^H \Sigma_{\mathbf{E}} \mathbf{a}_j}$$

[0086]

수학식 29

$$\text{SNR}_{y_j} = \text{SNR}_x \left(1 + \frac{\mathbf{a}_j^H \Sigma_{\mathbf{X}, \text{NG}} \mathbf{a}_j}{\mathbf{a}_j^H \text{diag}(\sigma_{x_1}^2, \dots, \sigma_{x_I}^2) \mathbf{a}_j} \right)$$

[0087]

[0088]이 수학식으로부터, 이 SNR은 신호 상관 매트릭스($\Sigma_{\mathbf{X}}$)의 대각선 및 비대각선 성분에 의존하는 항과의 승산에

의해 미리 정의된 SNR(SNR_x)로부터 얻어지는 것을 알 수 있다. 특히, $\Sigma_{\mathbf{X}, \mathbf{NG}}$ 가 제로 매트릭스가 되도록 신호($x_i(m)$)가 서로 비상관(uncorrelated)되면, 매트릭스화된 신호의 경험 SNR은 미리 정의된 SNR과 동일하고, 즉, $\Sigma_{\mathbf{X}, \mathbf{NG}} = \mathbf{0}_{I \times I}$ 이면, 모든 $j = 1, \dots, J$, 에 대하여,

수학식 30

$$\text{SNR}_{y_j} = \text{SNR}_x$$

[0089] SNR_{y_j} 는 $I \times I$ 행 및 열을 갖는 제로 매트릭스를 나타낸다. 즉, 신호($x_i(m)$)가 상관되면, 매트릭스화된 신호의 경험 SNR은 미리 정의된 SNR로부터 벗어날 수 있다. 최악의 경우, SNR_{y_j} 는 SNR_x 보다 매우 낮을 수 있다. 이 현상은 여기서 매트릭스화에서 잡음 언마스킹(noise unmasking)이라 한다.

[0090] 다음 섹션은 HOA에 대한 간략한 소개가 주어지고 처리될 신호(데이터 레이트 압축)을 정의한다.

[0091] HOA는 음원에서 자유로운 것으로 가정되는 콤팩트한 관심 영역 내의 음장(sound field)의 기술(description)에 기초한다. 이 경우, (구면 좌표 내의) 관심 영역 내의 위치($\mathbf{x} = [r, \theta, \phi]^T$) 및 시간(t)에서의 음압($p(t, x)$)의 시공간 작용은 동차 과동 방정식(homogeneous wave equation)에 의해 물리적으로 완전히 결정된다. 시간에 대하여 음압의 푸리에 변환, 즉,

수학식 31

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

[0092] 여기서, ω 는 각주파수를 나타내고, $\mathcal{F}_t \{ \cdot \}$ 는 $\int_{-\infty}^{\infty} p(t, \mathbf{x}) e^{-\omega t} dt$ 에 대응함 - 은 [10]에 따른 SH들(spherical harmonics)의 급수로 전개될 수 있음을 알 수 있다.

수학식 32

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

[0093] 수학식 32에서, c_s 는 음속을 나타내고 $k = \frac{\omega}{c_s}$ 는 각파수(angular wave number)이다. 또한, $j_n(\cdot)$ 은 제1종 및 차수 n 의 구면 베셀 함수(spherical Bessel function)을 나타내고, $Y_n^m(\cdot)$ 은 차수 n 및 디그리(degree) m 의 SH를 나타낸다.

[0094] 음장에 대한 완전한 정보는 실제로 음장 계수($A_n^m(k)$) 내에 포함된다.

[0095] SH는 일반적으로 복소수 함수들이라는 점에 유의해야 한다. 그러나, 이들의 적절한 선형 조합에 의해, 실수 함

수(real valued functions)를 얻고 이를 함수에 대하여 전개를 수행할 수 있다.

[0099] 수학식 32에서 압력 음장 기술(pressure sound field description)과 관련하여 음장은

수학식 33

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

[0100] [0101] 으로서 정의될 수 있고, 여기서, 음장 또는 진폭 밀도[9] $D(k c_s, \Omega)$ 는 각파수 및 각 방향 ($\Omega = [\theta, \phi]^T$)에 의존한다. 음장은 원거리 음장/근거리 음장, 불연속(discrete)/연속 소스로 구성될 수 있다[1]. 음장 계수(B_n^m)는 [1]에 의해 음장 계수(A_n^m)에 관련될 수 있다:

수학식 34

$$A_n^m = \begin{cases} 4\pi i^n B_n^m & \text{원거리 음장에 대하여} \\ -i k h_n^{(2)}(kr_s) B_n^m & \text{근거리 음장에 대하여}^1 \end{cases}$$

[0102] [0103] $-^1$ (e^{-ikr} 에 관련된) 진입 파에 대해 포지티브 주파수 및 제2종의 구면 항켈 함수($h_n^{(2)}$)를 이용함 -

[0104] 여기서, $h_n^{(2)}$ 는 제2종의 구면 항켈 함수(spherical Hankel function)이고 r_s 는 원점으로부터의 소스 거리이다.

[0105] HOA 도메인 내의 신호는 음장 또는 음장 계수의 역 푸리에 변환으로서 주파수 도메인 또는 시간 도메인에 표현될 수 있다. 다음의 설명은 유한 수의 음장 계수

수학식 35

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

[0106] [0107] 의 시간 도메인 표현의 사용을 가정할 것이다.

[0108] 수학식 33 내의 무한 급수는 $n=N$ 에서 절단(truncation)된다. 절단은 공간 벤드폭 제한에 대응한다. 계수(또는 HOA 채널)의 수는

수학식 36

$$O_{3D} = (N + 1)^2$$

[0109] [0110] (3D에 대하여)

[0111] 또는 2D만의 기술들(2D only descriptions)에 대하여 $O_{2D} = 2N + 1$ 로 주어진다. 계수(b_n^m)는 확성기에 의한 후속의 재생을 위해 하나의 시간 샘플(m)의 오디오 정보를 포함한다. 이들은 저장되거나 송신되고 따라서 데이터 레이트 압축의 대상이 된다.

[0112] 계수의 단일 시간 샘플(m)은 O_{3D} 엘리먼트를 갖는 벡터($b(m)$):

수학식 37

$$[0113] \quad b(m) := [b_0^0(m), b_1^{-1}(m), b_1^0(m), b_1^1(m), b_2^{-2}(m), \dots, b_N^N(m)]^T$$

[0114] 및 매트릭스(B)에 의한 M 개의 시간 샘플의 블록

수학식 38

$$[0115] \quad B := [b(m_{\text{START}} + 1), b(m_{\text{START}} + 2), \dots, b(m_{\text{START}} + M)]$$

[0116] 으로 표현될 수 있다.

[0117] 음장의 2차원 표현은 원형 조화 함수(circular harmonics)를 이용한 전개에 의해 도출될 수 있다. 이것은

$\theta = \frac{\pi}{2}$ 의 고정 경사, 계수의 상이한 가중 및 O_{2D} 계수($m = \pm n$)에 대한 감소된 세트를 이용하여 상기 제시된 일반적인 설명의 특수 경우로서 간주된다. 따라서, 다음의 모든 고려사항은 2D 표현에 적용되고, 구(sphere)라는 용어는 원(circle)이라는 용어로 대체될 필요가 있다.

[0118] 다음은 HOA 계수 도메인으로부터 공간 채널 기반 도메인으로의 변환 또는 그 반대를 설명한다. 수학식 33은 유닛 구상의 1개의 이산 공간 샘플 위치($\Omega_l = [\theta_l, \phi_l]^T$)에 대한 시간 도메인 HOA 계수를 이용하여 재기입될 수 있다:

수학식 39

$$[0119] \quad d_{\Omega_l} := \sum_{n=0}^N \sum_{m=-n}^n b_n^m Y_n^m(\Omega_l)$$

[0120] $L_{sd} = (N + 1)^2$ 구면 샘플 위치(Ω_l)를 가정하면, 이는 HOA 데이 블록(B)에 대한 벡터 표기로 재기입될 수 있고,

수학식 40

$$[0121] \quad W = \Psi_i B$$

[0122] 여기서, $\mathbf{W} = [\mathbf{w}(m_{\text{START}} + 1), \mathbf{w}(m_{\text{START}} + 2), \dots, \mathbf{w}(m_{\text{START}} + M)]^T$ 이고,

$\mathbf{w}(m) = [d_{\Omega_1}(m), \dots, d_{\Omega_{L_{sd}}}(m)]^T$ 는 L_{sd} 채널 신호의 단일 시간 샘플을 나타내고, 매트릭스

($\Psi_i = [\mathbf{y}_1, \dots, \mathbf{y}_{L_{sd}}]^H$)는 벡터($\mathbf{y}_l = [Y_0^0(\Omega_l), Y_1^{-1}(\Omega_l), \dots, Y_N^N(\Omega_l)]^T$)를 갖는다. 구면 샘플 위치가 매우 규칙적으로 선택되면,

수학식 41

$$\Psi_f \Psi_i = I$$

[0124] 를 갖는 매트릭스(Ψ_f)가 존재하고,

[0125] 여기서, I 는 $O_{3D} \times O_{3D}$ 아이덴티티 매트릭스이다. 수학식 40에 대한 해당 변환은

수학식 42

$$\mathbf{B} = \Psi_f \mathbf{W}$$

[0127] 에 의해 정의될 수 있다.

[0128] 수학식 42는 L_{sd} 구면 신호를 계수 도메인으로 변환하고 순방향 변환으로서 다음과 같이 재기입될 수 있고,

수학식 43

$$\mathbf{B} = DSHT\{\mathbf{W}\}$$

[0130] 여기서, DSHT{}는 이산 구면 조화 변환(Discrete Spherical Harmonics Transform)을 나타낸다. 해당 역 변환은 O_{3D} 계수 신호를 공간 도메인으로 변환하여 L_{sd} 채널 기반 신호를 형성하고 수학식 40은

수학식 44

$$\mathbf{W} = iDSHT\{\mathbf{B}\}$$

[0131] 이 된다.

[0133] 주어진 계수(B)로 시작하고 $\mathbf{B} = DSHT\{iDSHT\{\mathbf{B}\}\}$ 경우만 관심있기 때문에 이산 구면 조화 변환의 이 정의는 HOA 데이터의 데이터 레이트 압축에 관한 고려사항에 대하여 충분하다. 이산 구면 조화 변환의 더 엄격한 정의는 [2] 내에서 주어진다. DSHT에 대한 적절한 구면 샘플 위치 및 이러한 위치를 도출하는 절차는 [3], [4], [6], [5]에서 재검토될 수 있다. 샘플링 그리드의 예는 도 5에 도시된다.

[0134] 특히, 도 5는 블록(pE, pD)을 형성하는 인코더 및 디코더에서 사용되는 코드북에 대한 구면 샘플링 위치의 예,

즉, 도 5a)에서 $L_{sd} = 4$, 도 5b)에서 $L_{sd} = 9$, 도 5c)에서 $L_{sd} = 16$ 및 도 5d)에서 $L_{sd} = 25$ 을 나타낸다.

[0135] 다음에서, HOA 계수 데이터의 레이트 압축 및 잡음 언마스킹이 기재된다. 먼저, 테스트 신호가 정의되어 이하에서 사용되는 임의의 특성을 강조한다.

[0136] 방향(Ω_{s_1})에 위치하는 단일 원거리 음장 소스는 M개의 이산 시간 샘플의 벡터($\mathbf{g} = [g(m), \dots, g(M)]^T$)로 표현되고 인코딩에 의해 HOA 계수의 블록으로 표현되고:

수학식 45

$$\mathbf{B}_g = \mathbf{y} \mathbf{g}^T$$

[0138] 매트릭스(B_g)는 수학식 42와 유사하고 인코딩 벡터($\mathbf{y} = [Y_0^0(\Omega_{s_1}), Y_1^{-1}(\Omega_{s_1}), \dots, Y_N^N(\Omega_{s_1})]^T$)는

방향($\Omega_{s_1} = [\theta_{s_1}, \phi_{s_1}]^T$)에서 평가된 공액 복소수(conjugate complex) SH로 구성된다(실수 SH가 사용되면, 공액이 효과가 없다). 테스트 신호(B_g)는 HOA 신호의 가장 간단한 경우로서 간주될 수 있다. 더 많은 복소수 신호는 이러한 신호 중의 다수의 중첩(superposition)으로 구성된다.

[0139] HOA 채널의 직접 압축에 관하여, 다음은 HOA 계수 채널이 압축될 때 잡음 언마스킹이 발생하는 이유를 나타낸다. HOA 데이터(B)의 실제 블록의 \mathbf{O}_{3D} 계수 채널의 직접 압축 및 압축해제는 수학식 4와 유사한 코딩 잡음(E)을 유입할 것이다:

수학식 46

$$\widehat{\mathbf{B}} = \mathbf{B} + \mathbf{E}$$

[0141] 수학식 9에서처럼 상수(SNR_{B_g})를 가정한다. 이 신호를 확성기를 통해 리플레이하기 위하여 신호가 렌더링 될 필요가 있다. 이 프로세스는 다음과 같이 기술될 수 있고,

수학식 47

$$\widehat{\mathbf{W}} = \mathbf{A} \widehat{\mathbf{B}}$$

[0143] 여기서, 디코딩 매트릭스($\mathbf{A} \in \mathbb{C}^{L \times O_{3D}}$)(및 $\mathbf{A}^H = [\mathbf{a}_1, \dots, \mathbf{a}_L]$) 및 매트릭스($\widehat{\mathbf{W}} \in \mathbb{C}^{L \times M}$)는 L개의 스피커 신호의 M개의 시간 샘플을 유지한다. 이것은 수학식 14와 유사하다. 상술한 모든 고려사항을 적용하여, 스피커 채널(1)의 SNR은 (수학식 29과 유사한)

수학식 48

$$[0144] SNR_{w_l} = SNR_{B_g} \left(1 + \frac{\mathbf{a}_l^H \Sigma_{B,NG} \mathbf{a}_l}{\mathbf{a}_l^H \text{diag}(\sigma_{B_1}^2, \dots, \sigma_{B_{O_{3D}}}^2) \mathbf{a}_l} \right)$$

[0145] $\sigma_{B_0}^2$ 로 기술될 수 있고, 여기서, $\Sigma_{B,NG}$ 는 0번째 대각선 엘리먼트이고,

수학식 49

$$[0146] \Sigma_B = B B^H$$

[0147] 의 비대각선 엘리먼트를 유지한다.

[0148] 디코딩 매트릭스(A)가 영향을 받지 않음에 따라, 임의의 스피커 레이아웃으로 디코딩할 수 있어야 하기 때문에,

매트릭스($B=B_g$)는 대각선이 되어 $SNR_{w_l} = SNR_{B_g}$ 을 얻을 필요가 있다. 수학식 45 및 수학식 49로,

($B=B_g$) $\Sigma_B = \mathbf{y} \mathbf{g}^H \mathbf{g} \mathbf{y}^H = c \mathbf{y} \mathbf{y}^H$ 는 일정한 스칼라값 ($c = \mathbf{g}^T \mathbf{g}$)을 갖는 비대각선이 된다. SNR_{B_g} 와 비교하

여, 스피커 채널에서의 신호 대 잡음비(SNR_{w_l})는 감소한다. 그러나, 소스 신호(g) 또는 스피커 레이아웃 중의 어느 것도 인코딩 스테이지에서 통상 알려져 있지 않기 때문에, 계수 채널의 직접적인 손실 압축은 특히 낮은 데이터 레이트에 대하여 제어불가능한 언마스킹 효과로 이어질 수 있다.

[0149] 다음은 DSHT를 이용한 후에 HOA 계수가 공간 도메인에서 압축될 때 잡음 언마스킹이 발생하는 이유를 설명한다.

[0150] HOA 계수 데이터(B)의 현재 블록은 수학식 40에서 주어진 바와 같이 구면 조화 변환을 이용한 압축 전에 공간 도메인으로 변환되고,

수학식 50

$$[0151] W_{Sd} = \Psi_i B$$

[0152] 역 변환 매트릭스(Ψ_i)는 $L_{Sd} \geq 0_{3D}$ 공간 샘플 위치 및 공간 신호 매트릭스($W_{SH} \in \mathbb{C}^{L_{Sd} \times M}$)에 관련된다. 이들은 압축 및 압축 해제되고, 양자화 잡음에는 (수학식 4에 유사하게) 수학식 5에 따른 코딩 잡음 성분(E)이 부가된다:

수학식 51

$$[0153] \widehat{W}_{Sd} = W_{Sd} + E$$

[0154] 모든 공간 채널에 대해 상수인 $SNR(SNR_{Sd})$ 을 가정한다. 신호는 특성(41): $\Psi_f \Psi_i = I$ 를 갖는 변환 매트릭스

(Ψ_f)를 이용하여 계수 도메인 수학식 42로 변환된다. 계수($\widehat{\mathbf{B}}$)의 새로운 블록은

수학식 52

$$[0155] \quad \widehat{\mathbf{B}} = \Psi_f \widehat{\mathbf{W}}_{Sd}$$

[0156] 이 된다.

[0157] 이 신호는 디코딩 매트릭스(\mathbf{A}_D : $\widehat{\mathbf{W}} = \mathbf{A}_D \widehat{\mathbf{B}}$)를 적용함으로써 L개의 스피커 신호($\widehat{\mathbf{W}} \in \mathbb{C}^{L \times M}$)로 렌더링된다. 이는 수학식 52 및 $\mathbf{A} = \mathbf{A}_D \Psi_f$ 를 이용하여 재기입될 수 있다:

수학식 53

$$[0158] \quad \widehat{\mathbf{W}} = \mathbf{A} \widehat{\mathbf{W}}_{Sd}$$

[0159] 여기서, $\mathbf{A} \in \mathbb{C}^{L \times L_{Sd}}$ 를 갖는 혼합 매트릭스가 된다. 수학식 53은 수학식 14와 유사한 것으로 간주되어야 한다. 다시 상술한 모든 고려사항을 적용하여, 스피커 채널(1)의 SNR은 (수학식 29와 유사한)

수학식 54

$$[0160] \quad SNR_{w_l} = SNR_{s_d} \left(1 + \frac{\mathbf{a}_l^H \Sigma_{\mathbf{W}_{Sd,NG}} \mathbf{a}_l}{\mathbf{a}_l^H \text{diag}(\sigma_{S_{d_1}}^2, \dots, \sigma_{S_{d_{L_{Sd}}}}^2) \mathbf{a}_l} \right)$$

[0161] $\sigma_{S_{dl}}^2$ 로 기술될 수 있고, \mathbf{a}_l 는 1번째 대각선 엘리먼트이고 $\Sigma_{\mathbf{W}_{Sd,NG}}$ 는

수학식 55

$$[0162] \quad \Sigma_{\mathbf{W}_{Sd}} = \mathbf{W}_{Sd} \mathbf{W}_{Sd}^H$$

[0163] 의 비대각선 엘리먼트를 유지한다.

[0164] (임의의 확성기 레이아웃으로 렌더링할 수 있어야 하기 때문에) A_D 에 영향을 줄 방법이 없고 따라서 A 에 영향을 줄 방법이 없기 때문에, $\Sigma_{\mathbf{W}_{Sd}}$ 는 원하는 SNR을 유지하기 위하여 거의 대각선이 될 필요가 있다: 수학식 45로부터 간단한 테스트 신호를 이용하여 ($\mathbf{B} = \mathbf{B}_g$), $\Sigma_{\mathbf{W}_{Sd}}$ 는

수학식 56

$$\Sigma_{W_{Sd}} = c \Psi_i \mathbf{y} \mathbf{y}^H \Psi_i^H$$

[0165] [0166] 가 되고, $c = \mathbf{g}^T \mathbf{g}$ 는 상수이다. 고정된 구면 조화 변환(고정된 Ψ_i, Ψ_f)를 이용하면, $\Sigma_{W_{Sd}}$ 는 단지 매우

$$\frac{\mathbf{a}_l^H \Sigma_{W_{Sd},NG} \mathbf{a}_l}{\mathbf{a}_l^H \text{diag}\left(\sigma_{S_{d_1}}^2, \dots, \sigma_{S_{d_{L_{Sd}}}}^2\right) \mathbf{a}_l}$$

드문 경우에만 대각선이 되어 나빠질 수 있고, 상술한 바와 같이, 항($\Sigma_{W_{Sd},NG}$)은 계수 신호 공간 특성에 의존한다. 따라서, 구면 도메인 내의 HOA 계수의 낮은 레이트 손실 압축은 SNR 및 제어 불가능한 언마스킹 효과의 감소로 이어질 수 있다.

[0167] 본 발명의 기본 아이디어는 HOA 입력 신호의 공간 특성과 관련된 DSHT의 공간 샘플링 그리드의 회전 및 DSHT 자체로 구성되는 적응 DSHT(aDSHT)를 이용하여 잡음 언마스킹 효과를 최소화하는 것이다.

[0168] HOA 계수(O_{3D})(수학식 36)의 수에 매칭하는 다수의 구면 위치(L_{Sd})를 갖는 신호 적응 DSHT(aDSHT)가 이하에 기재된다. 먼저, 종래의 비적응 DSHT에서처럼 디폴트 구면 샘플 그리드가 선택된다. M개의 시간 샘플의 블록에 대하여, 구면 샘플 그리드는 항

수학식 57

$$\sum_{l=1}^{L_{Sd}} \sum_{j=1}^{L_{Sd}} |\Sigma_{W_{Sd},l,j}| - \sum \left(\sigma_{S_{d_1}}^2, \dots, \sigma_{S_{d_{L_{Sd}}}}^2 \right)$$

[0169]

[0170] 의 로그가 최소화되도록 회전되고, 여기서, $|\Sigma_{W_{Sd},l,j}|$ 는 (매트릭스 행 인덱스(1) 및 열 인덱스(j)를 갖는) $\Sigma_{W_{Sd}}$ 의 엘리먼트의 절대값이고 $\sigma_{S_{d_l}}^2$ 는 $\Sigma_{W_{Sd}}$ 의 대각선 엘리먼트이다. 이것은 수학식 54의 항

$$\frac{\mathbf{a}_l^H \Sigma_{W_{Sd},NG} \mathbf{a}_l}{\mathbf{a}_l^H \text{diag}\left(\sigma_{S_{d_1}}^2, \dots, \sigma_{S_{d_{L_{Sd}}}}^2\right) \mathbf{a}_l}$$

()을 최소화하는 것과 동일하다.

[0171] 가시화하면, 이 프로세스는 도 4에 도시된 바와 같이 단일 공간 샘플 위치가 가장 강한 소스 방향에 매칭하는 방식으로 DSHT의 구면 샘플링 그리드의 회전에 대응한다. 수학식 45($B=B_g$)로부터 간단한 테스트 신호를 이용하

여, 수학식 55의 항(W_{Sd})이 1을 제외하고 0과 근접한 모든 엘리먼트를 갖는 벡터($\in \mathbb{C}^{L_{Sd} \times 1}$)가 된다는 것을

알 수 있다. 결과적으로, $\Sigma_{W_{Sd}}$ 는 거의 대각선이 되고 원하는 SNR(SNR_{Sd})이 유지될 수 있다.

[0172] 도 4는 공간 도메인으로 변환된 테스트 신호(B_g)를 나타낸다. 도 4a)에서, 디폴트 샘플링 그리드가 사용되었고, 도 4b)에서, aDSHT의 회전된 그리드가 사용되었다. 공간 채널의 (dB 단위의) 관련된 $\Sigma_{W_{Sd}}$ 값은 대응하는 샘플 위치 주변의 보로노이(Voronoi) 셀의 칼라/그레이 변화로 도시된다. 공간 구조의 각각의 셀은 샘플링 포인트를 나타내고 셀의 밝기/어두움은 신호 강도를 나타낸다. 도 4b)에서 알 수 있는 바와 같이,

가장 강한 소스 방향을 찾고 샘플링 그리드는 사이드 중의 하나(즉, 단일 공간 샘플 위치)가 가장 강한 소스 방향에 매칭하도록 회전되었다. 이 사이드는 (강한 소스 방향에 대응한) 백색으로 도시되지만, 다른 사이드는 (낮은 소스 방향에 대응하여) 어둡다. 도 4a)에서, 즉, 회전 전에, 가장 강한 소스 방향에 매칭하지 않는 사이드는 없고, 몇 개의 사이드는 거의 그레이이고, 이는 각각의 샘플링 포인트에서 상당한(그러나 최대는 아닌) 강도의 오디오 신호가 수신되는 것을 의미한다.

[0173] 다음은 압축 인코더 및 디코더 내에서 사용되는 aDSHT의 주요 형성 블록을 설명한다.

[0174] 인코더 및 디코더 처리 형성 블록(pE 및 pD)의 세부사항은 도 6에 도시된다. 양 블록은 DSHT를 위한 기본인 구면 샘플링 위치 그리드의 동일 코드북을 소유한다. 초기에, 계수(\mathbf{O}_{3D})의 수는 공통 코드북에 따라 $L_{sd} = \mathbf{O}_{3D}$ 위치를 갖는 모듈(pE)에서 기본 그리드를 선택하는데 사용된다. L_{sd} 는 초기화를 위해 블록(pD)으로 송신되어 도 3에 지시된 바와 같이 동일한 기본 샘플링 위치 그리드를 선택해야 한다. 기본 샘플링 그리드는 매트릭스($\mathfrak{O}_{DSHT} = [\Omega_1, \dots, \Omega_{L_{sd}}]$)로 기술되고, 여기서, $\Omega_l = [\theta_l, \phi_l]^T$ 는 단위 구상의 위치를 정의한다. 상술한 바와 같이, 도 5는 기본 그리드의 예를 나타낸다.

[0175] 회전 탐색 블록(형성 블록 '최상의 회전 탐색')(320)으로의 입력은 계수 매트릭스(B)이다. 형성 블록은 수학식 57의 값이 최소가 되도록 기본 샘플링 그리드를 회전시키는 것을 책임진다. 회전은 '축-각(axis-angle)' 표현으로 표현되고, 이 회전과 관련된 압축된 축(ψ_{rot}) 및 회전 각(φ_{rot})은 사이드 정보(SI)로서 이 형성 블록으로 출력된다. 회전 축(ψ_{rot})은 원점으로부터 유닛 구상의 위치로의 단위 벡터로 기술될 수 있다. 구면 좌표에서, 이는 송신할 필요가 없는 1의 암시적 관련 반경을 갖는 2개의 각($\psi_{rot} = [\theta_{axis}, \phi_{axis}]^T$)로 표현될 수 있다. 3개의 각($\theta_{axis}, \phi_{axis}, \varphi_{rot}$)이 양자화되고 이전에 사용된 값의 재사용을 시그널링하여 사이드 정보(SI)를 생성하는 특수 탈출 패턴(special escape pattern)으로 엔트로피 코딩된다.

[0176] 형성 블록 ' Ψ_i 형성'(330)은 회전 축 및 각을 $\hat{\psi}_{rot}$ 및 $\hat{\varphi}_{rot}$ 으로 디코딩하고 이 회전을 기본 샘플링 그리드(\mathfrak{O}_{DSHT})에 적용하여 회전된 그리드($\hat{\mathfrak{O}}_{DSHT} = [\hat{\Omega}_1, \dots, \hat{\Omega}_{L_{sd}}]$)를 도출한다. 이는 벡터($\mathbf{y}_l = [Y_0^0(\hat{\Omega}_l), Y_1^{-1}(\hat{\Omega}_l), \dots, Y_N^N(\hat{\Omega}_l)]^T$)로부터 도출된 iDSHT 매트릭스($\Psi_i = [\mathbf{y}_1, \dots, \mathbf{y}_{L_{sd}}]$)를 출력한다.

[0177] 형성 블록 'iDSHT'(310)에서, HOA 계수 데이터(B)의 실제 블록은 $\mathbf{W}_{sd} = \Psi_i \mathbf{B}$ 에 의해 공간 도메인으로 변환된다.

[0178] 디코딩 처리 블록(pD)의 형성 블록 ' Ψ_f 형성'(350)은 회전 축 및 각을 수신하여 $\hat{\psi}_{rot}$ 및 $\hat{\varphi}_{rot}$ 으로 디코딩하고 이 회전을 기본 샘플링 그리드(\mathfrak{O}_{DSHT})에 적용하여 회전된 그리드($\hat{\mathfrak{O}}_{DSHT} = [\hat{\Omega}_1, \dots, \hat{\Omega}_{L_{sd}}]$)를 도출한다. iDSHT 매트릭스($\Psi_i = [\mathbf{y}_1, \dots, \mathbf{y}_{L_{sd}}]$)는 벡터($\mathbf{y}_l = [Y_0^0(\hat{\Omega}_l), Y_1^{-1}(\hat{\Omega}_l), \dots, Y_N^N(\hat{\Omega}_l)]^T$)로 도출되고 DSHT 매트릭스($\Psi_f = \Psi_i^{-1}$)는 디코딩 축상에서 산출된다.

[0179] 디코더 처리 블록(34) 내의 형성 블록 'DSHT'(340)에서, 공간 도메인 데이터($\widehat{\mathbf{W}}_{sd}$)의 실제 블록은 계수 도메인

데이터($\widehat{\mathbf{B}} = \Psi_f \widehat{\mathbf{W}}_{Sd}$)의 블록으로 다시 변환된다.

[0180] 다음에서, 압축 코덱의 전체 아키텍처를 포함하는 다양한 유리한 실시예가 기재된다. 제1 실시예는 단일 aDSHT를 사용한다. 제2 실시예는 스펙트럼 밴드에서 다수의 aDSHT를 사용한다.

[0181] 제1(기본) 실시예가 도 7에 도시된다. O_{3D} 계수 채널($b(m)$)의 인덱스(m)을 갖는 HOA 시간 샘플은 먼저 버퍼(71)에 저장되어 M 개의 샘플 및 시간 인덱스(μ)의 블록을 형성한다. $B(\mu)$ 는 상술한 바와 같이 형성 블록(pE)(72) 내의 적응 iDSHT를 이용하여 공간 도메인으로 변환된다. 공간 신호 블록($\mathbf{W}_{Sd}(\mu)$)은, AAC 또는 mp3 인코더 또는 단일 AAC 다채널 인코더(L_{Sd} 채널)처럼, L_{Sd} 오디오 압축 모노 인코더(73)에 입력된다. 비트스트림(S73)은 통합된 사이드 정보(SI) 또는 단일 다채널 비트스트림을 갖는 다수의 인코더 비트스트림 프레임의 멀티플렉싱된 프레임으로 구성되고, 사이드 정보(SI)는 보조 데이터로서 바람직하게 통합된다.

[0182] 각각의 압축 디코더 형성 블록은 일 실시예에서 비트스트림을 L_{Sd} 비트스트림 및 사이드 정보(SI)로 디멀티플렉싱하고(S73) 비트스트림을 L_{Sd} 모노 디코더로 공급하여 그들을 M 개의 샘플을 갖는 L_{Sd} 공간 오디오 채널로 디코딩하여 블록($\widehat{\mathbf{W}}_{Sd}(\mu)$)을 형성하고 $\widehat{\mathbf{W}}_{Sd}(\mu)$ 및 SI를 pD에 공급하는 디멀티플렉서(D1)를 포함한다. 비트스트림이 멀티플렉싱되지 않는 다른 실시예에서, 압축 디코더 형성 블록은, 비트스트림을 수신하여 그것을 L_{Sd} 다채널 신호($\widehat{\mathbf{W}}_{Sd}(\mu)$)로 디코딩하고 SI를 디팩킹(depack)하고 $\widehat{\mathbf{W}}_{Sd}(\mu)$ 및 SI를 pD에 공급하는 수신기(74)를 포함한다.

[0183] $\widehat{\mathbf{W}}_{Sd}(\mu)$ 는 디코더 처리 블록(pD)(75)에서 SI를 갖는 적응 DSHT를 이용하여 계수 도메인으로 변환되어 HOA 신호($B(\mu)$)의 블록을 형성하고, 이는 버퍼(76)에 저장되어 계수($b(m)$)의 시간 신호를 형성하도록 디프레이밍(deframe)된다.

[0184] 상술한 제1 실시예는 소정의 조건 하에서 2개의 결점을 가질 수 있다. 첫째로, 공간 신호 분배의 변경 때문에, 이전의 블록으로부터(즉, 블록(μ)으로부터 블록($\mu+1$)까지)의 블록킹 아티팩트(blocking artifact)이 존재할 수 있다. 두번째로, 동시에 1보다 많은 강한 신호가 존재할 수 있고 aDSHT의 역상관 효과는 꽤 작다.

[0185] 양 결점은 주파수 도메인에서 동작하는 제2 실시예에서 어드레스된다. aDSHT가 스케일 팩터 밴드 데이터에 적용되어 다수의 주파수 밴드 데이터를 조합한다. 블록킹 아티팩트는 OLA(overlay add) 처리로 TFT(Time to Frequency Transform)의 중첩 블록에 의해 회피된다. 개선된 신호 역상관은 SI_j 를 송신하기 위하여 데이터 레이트에서의 증가된 오버헤드를 희생하여 J개의 스펙트럼 밴드 내에서 본 발명을 사용함으로써 달성될 수 있다.

[0186] 도 9에 도시된 바와 같이, 제2 실시예의 임의의 더 많은 세부사항이 이하에 기재된다: 신호($b(m)$)의 각각의 계수 채널은 TFT(912)된다. 널리 사용되는 TFT에 대한 예는 MDCT(Modified Cosine Transform)이다. TFT 프레임 유닛(911)에서, 50% 중첩 데이터 블록(블록 인덱스(μ))이 구성된다. TFT 블록 변환 유닛(912)은 블록 변환을 수행한다. 스펙트럼 밴딩 유닛(913)에서, TFT 주파수 밴드가 결합되어 J개의 새로운 스펙트럼 밴드 및 관련

신호($B_j(\mu)$)($\in \mathbb{C}^{0_{3D} \times K_j}$)를 형성하고, 여기서, K_j 는 밴드(j) 내의 주파수 계수의 수를 나타낸다. 이들 스펙트럼 밴드는 복수의 처리 블록(914)에서 처리된다. 이들 스펙트럼 밴드의 각각에 대하여 신호

($\mathbf{W}_{j_{Sd}}(\mu) \in \mathbb{C}^{L_{Sd} \times K_j}$) 및 사이드 정보(SI_j)를 생성하는 하나의 처리 블록(pE_j)이 존재한다. 스펙트럼 밴드는 (AAC/mp3 스케일 팩터 밴드처럼) 손실있는 오디오 압축 방법의 스펙트럼 밴드와 매칭하거나 매우 거친 그래뉼러리티를 가질 수 있다. 후자의 경우, TFT 블록이 없이 채널 독립 손실 오디오 압축(915)은 밴딩(banding)을 재배치할 필요가 있다. 처리 블록(914)은 각각의 오디오 채널에 일정한 비트 레이트를 할당하는 주파수 도메인

내의 L_{Sd} 다채널 오디오 인코더처럼 동작한다. 비트스트림은 비트스트림 팩킹 블록(916)에서 포맷화된다.

[0187] 디코더는 비트스트림(적어도 그 일부)를 수신하거나 저장하고 이를 디팩킹(921)하여 TFT 없이 채널 독립 오디오 디코딩을 위해 오디오 데이터를 다채널 오디오 디코더(922)에 공급하고 사이드 정보(SI_j)를 복수의 디코딩 처리 블록(pD_j)(923)에 공급한다. TFT 없이 채널 독립 오디오 디코딩을 위한 오디오 디코더(922)는 오디오 정보를 디

코딩하고 디코딩 처리 블록(pD_j)(923)으로의 입력으로서 J개의 스펙트럼 밴드 신호($\widehat{\mathbf{W}}_{Sd}(\mu)$)를

포맷화하고, 이들 신호는 HOA 계수 도메인으로 변환되어 $\widehat{\mathbf{B}}_j(\mu)$ 를 형성한다. 스펙트럼 디밴딩 블록(924)에서, J개의 스펙트럼 밴드는 재그룹화되어 TFT의 밴딩에 매칭한다. 이들은 iTFT 및 OLA 블록(925)에서 시간 도메인으로 변환되고, 이는 블록 중첩 OLA(overlay add) 처리를 이용한다. 마지막으로, iTFT 및 OLA 블록(925)의 출력은 TFT 디프레이밍 블록(926)에서 디프레이밍되어 신호($\widehat{\mathbf{b}}(m)$)를 생성한다.

[0188] 본 발명은 SNR 증가가 채널 간의 상호 상관으로부터 기인한다는 결과에 기초한다. 지각적인 코더는 단지 각각의 개별 단일 채널 신호 내에 발생하는 코딩 잡음 마스킹 효과를 고려한다. 그러나, 이러한 효과는 일반적으로 비선형이다. 따라서, 이러한 단일 채널을 새로운 신호로 매트릭스화하면, 잡음 언마스킹이 발생할 가능성이 있다. 이것은 매트릭스화 동작 후에 통상 코딩 잡음이 증가하는 이유이다.

[0189] 본 발명은 원치 않는 잡음 언마스킹 효과를 최소화하는 적응 DSHT에 의해 채널의 역상관을 제안한다. aDSHT는 압축 코더 및 디코더 아키텍쳐 내에서 통합된다. 이는 DSHT의 공간 샘플링 그리드를 HOA 입력 신호의 공간 특성으로 조절하는 회전 동작을 포함하기 때문에 적응적이다. aDSHT는 적응 회전 및 실제 종래의 DSHT를 포함한다. 실제 DSHT는 종래 기술에 기재된 바와 같이 구성될 수 있는 매트릭스이다. 적응 회전이 매트릭스에 적용되고, 이는 채널간 상관의 최소화로 이어지고, 그러므로, 매트릭스화후 SNR 증가의 최소화로 이어진다. 회전 축 및 각은 분석적이지 않게 자동화된 탐색 동작에 의해 탐색된다. 회전 축 및 각은 디코딩 후 및 매트릭스화 전의 재상관을 가능하게 하기 위하여 인코딩 및 송신되고, 역 적응 DSHT(iADSHT)가 사용된다.

[0190] 일 실시예에서, TTF 및 스펙트럼 밴딩이 수행되고 aDSHT/iADSHT가 독립적으로 각각의 스펙트럼 밴드에 적용된다.

[0191] 도 8a)는 본 발명의 일 실시예에서 잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 방법의 플로우챠트를 나타낸다. 도 8b)는 본 발명의 일 실시예에서 잡음 감소를 위한 다채널 HOA 오디오 신호를 디코딩하는 방법의 플로우챠트를 나타낸다.

[0192] 도 8a)에 도시된 실시예에서, 잡음 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 방법은 역 적응 DSHT를 이용하여 채널을 역상관하는 단계(81) - 역 적응 DSHT는 회전 동작 및 역 DSHT(812)를 포함하고, 회전 동작은 iDSHT의 공간 샘플링 그리드를 회전함(811) -, 역상관된 채널의 각각을 지각적으로 인코딩하는 단계(82), (사이드 정보(SI)로서) 회전 정보를 인코딩하는 단계(83) - 회전 정보는 상기 회전 동작을 정의하는 파라미터를 포함함 -, 및 지각적으로 인코딩된 오디오 채널 및 인코딩된 회전 정보를 송신 또는 저장하는 단계(84)를 포함한다.

[0193] 일 실시예에서, 역 적응 DSHT는 초기 디폴트 구면 샘플 그리드를 선택하는 단계, 가장 강한 소스 방향을 결정하는 단계, M개의 시간 샘플의 블록에 대하여 구면 샘플 그리드를 회전하여 단일 공간 샘플 위치가 가장 강한 소스 방향에 매칭하도록 하는 단계를 포함한다.

[0194] 일 실시예에서, 구면 샘플 그리드는 항:

$$\sum_{l=1}^{L_{Sd}} \sum_{j=1}^{L_{Sd}} | \Sigma_{W_{Sd}, l,j} | - \sum \left(\sigma_{S_d 1}^2, \dots, \sigma_{S_d L_{Sd}}^2 \right)$$

[0195] 의 로그가 최소화되도록 회전되고, 여기서, $| \Sigma_{W_{Sd}, l,j} |$ 는 (매트릭스 행 인덱스(l) 및 열 인덱스(j)를 갖는)

$\Sigma_{W_{Sd}}$ 의 엘리먼트의 절대값이고 $\sigma_{S_d l}^2$ 는 $\Sigma_{W_{Sd}}$ 의 대각선 엘리먼트이고, 여기서, $\Sigma_{W_{Sd}} = \mathbf{W}_{Sd} \mathbf{W}_{Sd}^H$

이고 \mathbf{W}_{Sd} 는 오디오 채널의 수×블록 처리 샘플의 수 매트릭스이고, \mathbf{W}_{Sd} 는 aDSHT의 결과이다.

[0197] 도 8b)에 도시된 실시예에서, 감소된 잡음을 갖는 코딩된 다채널 HOA 오디오 신호를 디코딩하는 방법은 (사이드 정보(SI) 내의) 다채널 HOA 오디오 신호 및 채널 회전 정보를 수신하는 단계(85), 수신된 데이터를 압축 해제하는 단계(86) - 지각적 디코딩이 사용됨 -, 적응 DSHT를 이용하여 각 채널을 공간적으로 디코딩하는 단계(87) - DSHT(872) 및 상기 회전 정보에 따른 DSHT의 공간 샘플링 그리드의 회전(871)이 수행되고 지각적으로 디코딩된 채널이 역상관됨 -, 및 지각적으로 디코딩되고 역상관된 채널을 매트릭스화하는 단계(88) - 확성기 위치에 맵핑된 재생가능 오디오 신호가 얻어짐 - 를 포함한다.

[0198] 일 실시예에서, 적응 DSHT는 적응 DSHT에 대한 초기 디폴트 구면 샘플 그리드를 선택하는 단계 및 M개의 시간 샘플의 블록에 대하여 상기 회전 정보에 따라 구면 샘플 그리드를 회전하는 단계를 포함한다.

[0199] 일 실시예에서, 회전 정보는 3개의 성분을 갖는 공간 벡터($\hat{\boldsymbol{\psi}}_{rot}$)이다. 회전 축($\boldsymbol{\psi}_{rot}$)은 단위 벡터로 기술될 수 있음에 주의해야 한다.

[0200] 일 실시예에서, 회전 정보는 3개의 각, 즉, $\theta_{axis}, \phi_{axis}, \varphi_{rot}$ 으로 구성된 벡터이고, 여기서, $\theta_{axis}, \phi_{axis}$ 는 구면 좌표 내의 1의 암시적 반경을 갖는 회전 축에 대한 정보를 정의하고, φ_{rot} 는 이 축 주변의 회전 각을 정의한다.

[0201] 일 실시예에서, 작은 양자화되고 사이드 정보(SI)를 생성하는 이전의 값의 재사용을 시그널링하는(즉, 지시하는) 탈출 패턴(escape pattern)(즉, 전용 비트 패턴)으로 엔트로피 코딩된다.

[0202] 일 실시예에서, 잡음을 감소를 위한 다채널 HOA 오디오 신호를 인코딩하는 장치는 역 적응 DSHT를 이용하여 채널을 역상관하는 역상관기 - 역 적응 DSHT는 회전 동작 및 역 DSHT(iDSHT)를 포함하고, 회전 동작은 iDSHT의 공간 샘플링 그리드를 회전함 -, 역상관된 채널의 각각을 지각적으로 인코딩하는 지각적 인코더, 회전 정보를 인코딩하는 사이드 정보 인코더 - 회전 정보는 상기 회전 동작을 정의하는 파라미터를 포함함 -, 및 지각적으로 인코딩된 오디오 채널 및 인코딩된 회전 정보를 송신 또는 저장하는 인터페이스를 포함한다.

[0203] 일 실시예에서, 감소된 잡음을 갖는 다채널 HOA 오디오 신호를 디코딩하는 장치는 인코딩된 다채널 HOA 오디오 신호 및 채널 회전 정보를 수신하는 인터페이스 수단(330), 각각의 채널을 지각적으로 디코딩하는 지각적 디코더를 이용하여 수신된 데이터를 압축 해제하는 압축 해제 모듈(33), 지각적으로 디코딩된 채널을 재상관하는 상관기(34) - DSHT 및 상기 회전 정보에 따른 DSHT의 공간 샘플링 그리드의 회전이 수행됨 -, 및 지각적으로 디코딩되고 상관된 채널을 매트릭스화하는 럭셔리(MX) - 확성기 위치에 맵핑된 재생가능 오디오 신호가 얻어짐 - 를 포함한다. 원리적으로, 상관기(34)는 공간 디코더로서 동작한다.

[0204] 일 실시예에서, 감소된 잡음을 갖는 다채널 HOA 오디오 신호를 디코딩하는 장치는 인코딩된 다채널 HOA 오디오 신호 및 채널 회전 정보를 수신하는 인터페이스 수단(330), 각각의 채널을 지각적으로 디코딩하는 지각적 디코더로 수신된 데이터를 압축 해제하는 압축 해제 모듈(33), aDSHT를 이용하여 지각적으로 디코딩된 채널을 상관하는 상관기(34) - DSHT 및 상기 회전 정보에 따른 DSHT의 공간 샘플링 그리드의 회전이 수행됨 -, 및 지각적으로 디코딩되고 상관된 채널을 매트릭스화하는 럭셔리(MX) - 확성기 위치에 맵핑된 재생가능 오디오 신호가 얻어짐 - 를 포함한다.

[0205] 일 실시예에서, 디코딩 장치 내의 적응 DSHT는 적응 DSHT에 대한 초기 디폴트 구면 샘플 그리드를 선택하는 수단, M개의 시간 샘플의 블록에 대하여 상기 회전 정보에 따라 디폴트 구면 샘플 그리드를 회전하는 회전 처리 수단, 및 회전된 구면 샘플 그리드에 대하여 DSHT를 수행하는 변환 처리 수단을 포함한다.

[0206] 일 실시예에서, 디코딩 장치 내의 상관기(34)는 적응 DSHT를 이용하여 각 채널을 동시에 공간적으로 디코딩하는 복수의 공간 디코딩 유닛(922)를 포함하고, 스펙트럼 디밴딩을 수행하는 스펙트럼 디밴딩 유닛(924) 및 OLA(overlay Add) 처리로 역 TFT를 수행하는 iTFT 및 OLA 유닛(925)를 더 포함하고, 스펙트럼 디밴딩 유닛은 iTFT 및 OLA 유닛에 그 출력을 제공한다.

[0207] 모든 실시예에서, 감소된 잡음이라는 용어는 적어도 코딩 잡음 언마스킹의 회피에 관한 것이다.

[0208] 오디오 신호의 지각적 코딩은 오디오의 인간 지각에 적응되는 코딩을 의미한다. 오디오 신호를 지각적으로 코딩할 때, 양자화는 통상 베이스밴드 오디오 신호 샘플에 대하여 수행되지 않고 오히려 인간 지각에 관련된 개별 주파수 밴드에서 수행된다는 것에 유의해야 한다. 그러므로, 신호 전력 및 양자화 잡음 간의 비는 개별 주파수 밴드마다 변할 수 있다. 그러므로, 지각적 코딩은 통상 리던던시 및/또는 무관(irrelevancy) 정보의 감소를 포함하지만, 공간 코딩은 통상 채널 간의 공간 관계에 관한 것이다.

[0209] 상술한 기술은 KLT(Karhunen-Loeve-Transformation)를 이용하는 역상관에 대한 대안으로서 간주될 수 있다. 본 발명의 하나의 이점은 딱 3개의 각을 포함하는 사이드 정보의 양의 강한 감소이다. KLT는 사이드 정보로서 블록 상관 매트릭스의 계수를 요구하고, 따라서, 상당히 많은 데이터를 요구한다. 또한, 여기에 개시된 기술은 다음의 처리 블록으로 진행할 때 천이 아티팩트(transition artifact)를 감소시키기 위하여 비틀기(tweaking) (또는 미세 조정(fine-tuning))을 허용한다. 이것은 후속의 지각적 코딩의 압축 품질에 유리하다.

[0210] 표 1은 aDSHT 및 KLT 간의 직접 비교를 제공한다. 약간의 유사성이 존재하지만, aDSHT는 KLT보다 상당한 이점을 제공한다.

표 1

aDSHT 대 KLT의 비교

	sDSHT	KLT
정의	B는 $(N+1)^2$ 행(계수) 및 T 열(시간 샘플)을 갖는 N차 HOA 신호 매트릭스; W는 $(N+1)^2$ 행(채널) 및 T 열(시간 샘플)을 갖는 공간 매트릭스	
인코더, 공간 변환	역 aDSHT $\mathbf{W}_{sd} = \Psi_i \mathbf{B}$	KLT $\mathbf{W}_k = \mathbf{K} \mathbf{B}$
변환 매트릭스	인코더 및 디코더에 알려진 $(N+1)^2$ 구면 샘플 위치를 갖는 규칙적인 구면 샘플링 그리드가 선택된다. 이 그리드는 축 Ψ_{rot} 및 회전각(Φ_{rot}) 주위를 회전하고, 이는 이전에 도출되었다(이하의 리마크 참조). 그 그리드의 모드 매트릭스(Ψ_f)가 생성된다(즉, 이를 위치의 구면 조화): $\Psi_i = \Psi_f^{-1}$ (또는 공간 채널의 수가 $(N+1)^2$ 보다 커지면 $\Psi_f \Psi_i = I$ 를 갖는 더 일반적인 $\Psi_i = \Psi_f^+$). 변환 매트릭스는 회전된 구면 그리드의 혼모드 매트릭스이다. 회전은 신호 구동되고 처리 블록마다 업데이트된다.	공분산 매트릭스: $\mathbf{C} = \mathbf{B} \mathbf{B}^H$ 를 형성. 고유치 분해: $\mathbf{C} = \mathbf{K}^H \Lambda \mathbf{K}$, A의 대각선의 고유치 및 \mathbf{K}^H 내에 배치된 관련된 고유 벡터를 갖고, 임의의 직교 변환 내에서 처럼 $\mathbf{K} \mathbf{K}^H = 1$ 를 갖는다. 변환 매트릭스는 처리 블록마다 신호(B)로부터 도출된다.
송신될 사이드 정보	축(Ψ_{rot}) 및 회전각(Φ_{rot})이 예를 들어 3개의 값 $(\theta_{axis}, \phi_{axis}, \Psi_{rot})$ 으로서 코딩된다.	C의 엘리먼트의 절반보다 많음(즉, $\frac{(N+1)^4 + (N+1)^2}{2}$ 값) 또는 K(즉, $(N+1)^4$ 값)
손실있는 분해 공간 신호	공간 신호는 손실 코딩된다(코딩 잡음 (E_{cod})). T개의 샘플의 블록은 $\widehat{\mathbf{W}}_{sd}$ 로서 배치된다.	공간 신호는 손실 코딩된다(코딩 잡음 ($\widehat{\mathbf{E}}_{cod}$)). T개의 샘플들의 블록은 $\widehat{\mathbf{W}}_k$ 로서 배치된다.
디코더, 역 공간 변환	$\widehat{\mathbf{B}} = \Psi_f \widehat{\mathbf{W}}_{sd} = \mathbf{B} + \Psi_f E_{cod}$	$\widehat{\mathbf{B}}_k = K \widehat{\mathbf{W}}_k = \mathbf{B} + K \widehat{\mathbf{E}}_{cod}$
리마크	일 실시예에서, 그리드는 샘플링 위치가 B 내의 가장 강한 신호 방향에 매칭하도록 회전한다. KLT에 대하여 이용가능한 것처럼, 공분산 매트릭스의 분석이 사용될 수 있다. 실제로, 더 간단하고 계산적으로 덜 복잡하기 때문에, 블록마다 원활하게 회전을 적용/변경하도록 하는 신호 트랙킹 모델이 이용될 수 있고, 이는 손실있는 (지각적인) 코딩 블록 내의 블록킹 아티팩트의 생성을 피한다.	

[0212] 본 발명의 기본적인 특징이 바람직한 실시예에 적용되는 것으로 도시되고, 기재되고 지시되지만, 개시된 장치의 형태 및 세부 사항 및 그 동작에 있어서 기재된 장치 및 방법의 다양한 생략 및 대체 및 변경이 본 발명의 사상을 벗어나지 않고 통상의 기술자에 의해 가능하다. 실질적으로 동일한 방식으로 실질적으로 동일한 기능을 수행하여 동일한 결과를 달성하는 이를 엘리먼트의 모든 조합이 본 발명의 범위 내에 있다. 기재된 하나의 실시예로부터 다른 실시예로의 엘리먼트의 대체가 또한 의도되고 고려된다. 본 발명은 예로서 단순히 기재되고 본 발명의 범위를 벗어나지 않고 세부사항의 변경이 가능함을 이해할 것이다.

[0213] 설명 및 (적절하다면) 청구범위 및 도면에 개시된 각각의 특징은 독립적으로 또는 임의의 적절한 조합으로 제공될 수 있다.

[0214] 적절하다면, 특징은 하드웨어, 소프트웨어 또는 그 조합으로 구현될 수 있다. 적용가능하다면, 접속은 무선 접속 또는 유선 접속으로 구현될 수 있고, 반드시 직접이거나 전용은 아닐 수 있다.

[0215] 청구범위에 나타내는 참조 번호는 단지 예시적인 것으로 청구범위의 범위에 대한 제한적 효과를 갖지 않는다.

[0216]

인용 문헌

- [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.
- [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.
- [3] Jörg Fliege. Integration nodes for the sphere, <http://www.personal.soton.ac.uk/jf1w07/nodes/nodes.html>
- [4] Jörg Fliege and Ulrike Maier. A two-stage approach for computing cubature formulae for the sphere. Technical Report, Fachbereich Mathematik, Universität Dortmund, 1999.
- [5] R. H. Hardin and N. J. A. Sloane. Webpage: Spherical designs, spherical t-designs. <http://www2.research.att.com/~njas/sphdesigns>
- [6] R. H. Hardin and N. J. A. Sloane. McLaren's improved snub cube and other new spherical designs in three dimensions. *Discrete and Computational Geometry*, 15:429–441, 1996.
- [7] Erik Hellerud, Ian Burnett, Audun Solvang, and U. Peter Svensson. Encoding higher order Ambisonics with AAC. In 124th AES Convention, Amsterdam, May 2008.
- [8] Peter Jax, Jan-Mark Batke, Johannes Boehm, and Sven Kordon. Perceptual coding of HOA signals in spatial domain. European patent application EP2469741A1 (PD100051).
- [9] Boaz Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J. Acoust. Soc. Am.*, 4(116):2149–2157, October 2004.
- [10] Earl G. Williams. Fourier Acoustics, volume 93 of *Applied Mathematical Sciences*. Academic Press, 1999.

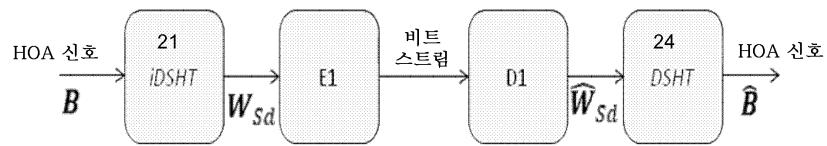
[0217]

도면

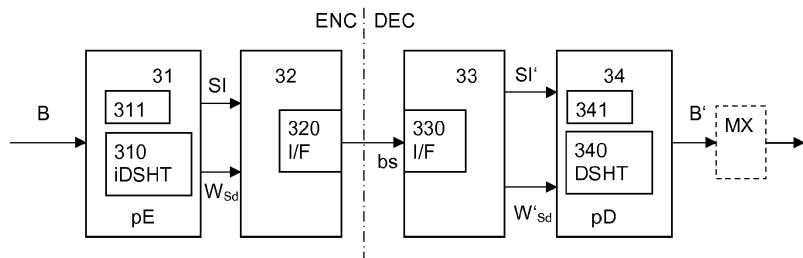
도면1



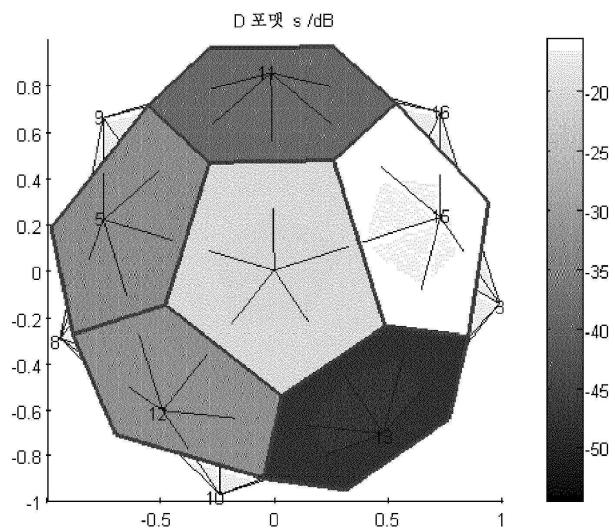
도면2



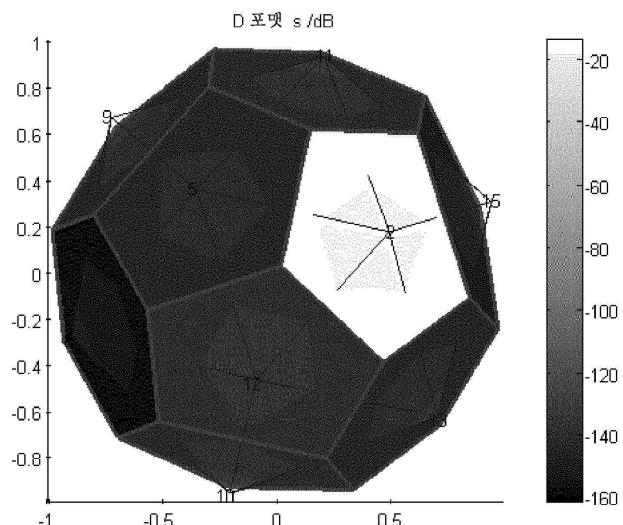
도면3



도면4

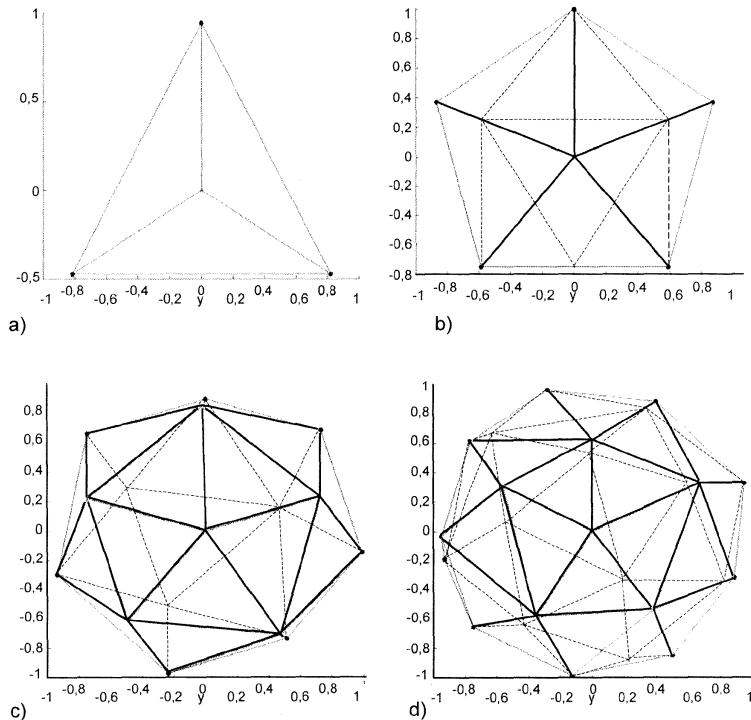


a)

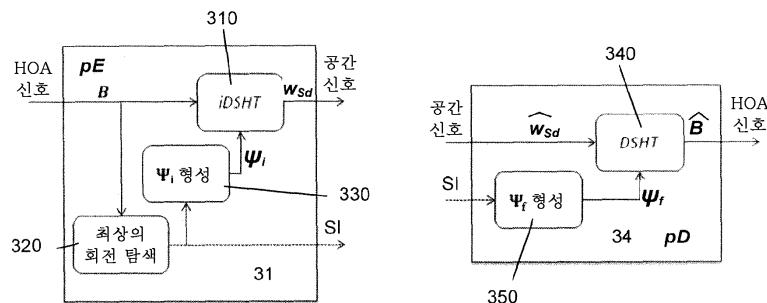


b)

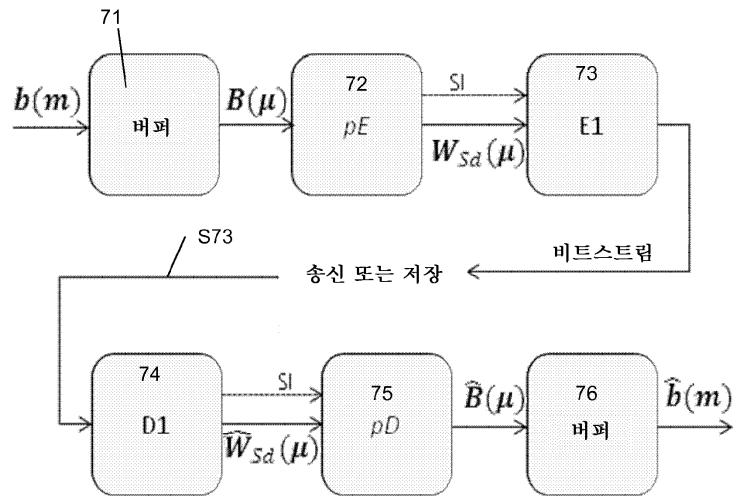
도면5



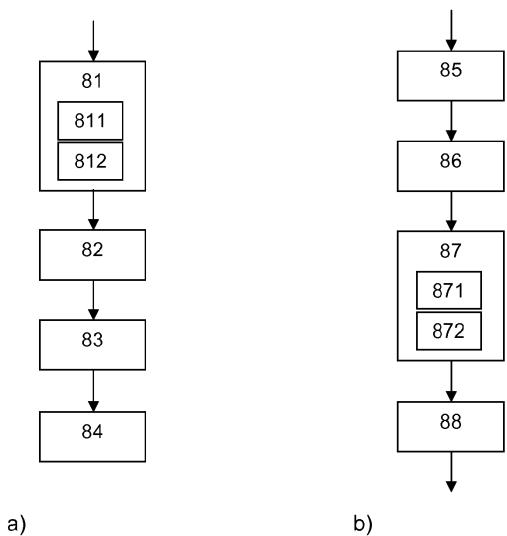
도면6



도면7



도면8



도면9

