(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0078465
(43) 공개일자 2025년06월02일

(51) 국제특허분류(Int. Cl.)
*G10L 19/008* (2014.01) *G10L 19/16* (2013.01)
(52) CPC특허분류
*G10L 19/008* (2020.08)
*G10L 19/167* (2013.01)
(21) 출원번호 10-2025-7011149
(22) 출원일자(국제) 2023년09월15일
심사청구일자 없음
(85) 번역문제출일자 2025년04월04일
(86) 국제출원번호 PCT/EP2023/075436
(87) 국제공개번호 WO 2024/074284
국제공개일자 2024년04월11일
(30) 우선권주장
63/378,500 2022년10월05일 미국(US)
63/534,458 2023년08월24일 미국(US)

(71) 출원인
**돌비 인터네셔널 에이비**
아일랜드 디02 브이케이60 더블린 그랜드 커널 독 랜즈 블록 씨 서 존 로저슨스 키 77
(72) 발명자
**먼드트, 하랄드**
미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레버러토리즈 인크. 내
**펀하이젠, 헤이코**
미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레버러토리즈 인크. 내
*(뒷면에 계속)*
(74) 대리인
**특허법인 광장리앤고**

전체 청구항 수 : 총 26 항

(54) 발명의 명칭 **오디오 비트스트림의 효율적인 인코딩 및 디코딩을 위한 방법, 장치 및 매체**

## (57) 요 약

복수의 오디오 신호를 포함하는 오디오 프로그램의 인코딩된 비트스트림의 프레임을 생성하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은, 복수의 오디오 신호 각각에 대해, 개개의 오디오 신호가 연관된 재생 디바이스를 나타내는 정보를 수신하는 단계, 하나 이상의 인코딩된 오디오 신호를 획득하기 위해 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계, 개개의 재생 디바이스와 연관된 하나 이상의 인코딩된 오디오 신호를 프레임의 제1 독립적인 블록에 결합하는 단계, 복수의 오디오 신호 중 하나 이상의 다른 오디오 신호를 하나 이상의 추가적인 독립적인 블록으로 인코딩하는 단계, 및 제1 독립적인 블록 및 하나 이상의 추가적인 독립적인 블록을 인코딩된 비트스트림의 프레임에 결합하는 단계를 포함한다.

## 대 표 도

<table>
<tr><td></td><td colspan="3">프레임 N</td><td></td></tr>
<tr><td></td><td>블록 ID=0,<br>F=N, P=0</td><td>블록 ID=1,<br>F=N, P=0</td><td>블록 ID=2,<br>F=N, P=0</td><td></td></tr>
<tr><td>0110010101</td><td>10011110101001001</td><td>001100101010101</td><td>00101010011011010010101</td><td>0100111010</td></tr>
</table>

(72) 발명자

**솔링, 크리스토퍼**

미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레버러토리즈 인크. 내

**슈그, 마이클**

미국 캘리포니아 94103 샌프란시스코 마켓 스트리트 1275 돌비 레버러토리즈 인크. 내

# 명 세 서

## 청구범위

### 청구항 1

복수의 오디오 신호를 포함하는 오디오 프로그램의 인코딩된 비트스트림(encoded bitstream)의 프레임을 생성하기 위한 방법으로서, 상기 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 상기 방법은:

상기 복수의 오디오 신호 각각에 대해, 개개의 오디오 신호가 연관된 재생(playback) 디바이스를 표시하는 정보를 수신하는 단계;

하나 이상의 인코딩된 오디오 신호를 획득하기 위해 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계;

상기 개개의 재생 디바이스와 연관된 상기 하나 이상의 인코딩된 오디오 신호를 상기 프레임의 제1 독립적인 블록에 결합하는 단계;

상기 복수의 오디오 신호 중 하나 이상의 다른 오디오 신호를 하나 이상의 추가적인 독립적인 블록으로 인코딩하는 단계; 및

상기 제1 독립적인 블록 및 상기 하나 이상의 추가적인 독립적인 블록을 상기 인코딩된 비트스트림의 상기 프레임에 결합하는 단계를 포함하는, 방법.

### 청구항 2

제1항에 있어서, 둘 이상의 오디오 신호는 상기 재생 디바이스와 연관되고, 상기 둘 이상의 오디오 신호 각각은 상기 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된(bandlimited) 신호이고, 상기 대역제한된 신호 각각에 대해 상이한 인코딩 기술이 사용되는, 방법.

### 청구항 3

제2항에 있어서, 상기 대역제한된 신호 각각에 대해 상이한 심리음향 모델(psychoacoustic model) 및/또는 상이한 비트 할당 기술이 사용되는, 방법.

### 청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 인코딩된 신호의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델(buffer fullness model)에 의해 제약되는, 방법.

### 청구항 5

제1항에 있어서, 상기 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계는 상기 개개의 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 상기 프레임의 상기 제1 독립적인 블록으로 공동으로 인코딩하는 단계를 포함하는, 방법.

### 청구항 6

제5항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 이상의 오디오 신호에 걸쳐 하나 이상의 스케일 팩터(scale factor)를 공유하는 단계를 포함하는, 방법.

### 청구항 7

제6항에 있어서, 상기 둘 이상의 오디오 신호는 공간적으로 관련되는, 방법.

### 청구항 8

제7항에 있어서, 상기 둘 이상의 공간적으로 관련된 오디오 신호는 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 또는 우측 정상 채널을 포함하는, 방법.

**청구항 9**

제5항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 커플링 툴을 적용하는 단계를 포함하며, 상기 커플링 툴을 적용하는 단계는:

둘 이상의 오디오 신호를 명시된 주파수 위의 복합 신호로 결합하는 단계; 및

상기 둘 이상의 오디오 신호 각각에 대해, 상기 복합 신호의 에너지 및 각각의 개개의 신호의 에너지에 관한 스케일 팩터를 결정하는 단계를 포함하는, 방법.

**청구항 10**

제5항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 초과의 신호에 공동-코딩 툴(joint-coding tool)을 적용하는 단계를 포함하는, 방법.

**청구항 11**

인코딩된 비트스트림의 프레임으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하기 위한 방법으로서, 상기 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 상기 방법은:

상기 인코딩된 비트스트림으로부터, 상기 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 독립적인 블록을 식별하는 단계;

상기 인코딩된 비트스트림으로부터, 상기 인코딩된 데이터의 식별된 독립적인 블록을 추출하는 단계;

하나 이상의 디코딩된 오디오 신호를 획득하기 위해 상기 인코딩된 데이터의 독립적인 블록으로부터 상기 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호를 디코딩하는 단계;

상기 인코딩된 비트스트림으로부터, 하나 이상의 추가적인 오디오 신호에 대응하는 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 식별하는 단계; 및

상기 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 디코딩하거나 스킵하는 단계를 포함하는, 방법.

**청구항 12**

제11항에 있어서, 둘 이상의 오디오 신호는 상기 재생 디바이스와 연관되고, 상기 둘 이상의 오디오 신호 각각은 상기 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된 신호이고, 상이한 디코딩 기술이 상기 둘 이상의 오디오 신호를 디코딩하는 데 사용되는, 방법.

**청구항 13**

제12항에 있어서, 상이한 심리음향 모델 및/또는 상이한 비트 할당 기술이 상기 대역제한된 신호 각각을 인코딩하기 위해 사용된, 방법.

**청구항 14**

제12항 또는 제13항에 있어서, 상기 인코딩된 비트스트림의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델에 의해 제약되는, 방법.

**청구항 15**

제11항에 있어서, 상기 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호를 디코딩하는 단계는 개개의 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 상기 인코딩된 데이터의 독립적인 블록으로부터 공동으로 디코딩하는 단계를 포함하는, 방법.

**청구항 16**

제15항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는

단계는 둘 이상의 오디오 신호에 걸쳐 공유되는 스케일 팩터를 추출하는 단계를 포함하는, 방법.

**청구항 17**

제16항에 있어서, 상기 둘 이상의 오디오 신호는 공간적으로 관련되는, 방법.

**청구항 18**

제17항에 있어서, 상기 둘 이상의 공간적으로 관련된 오디오 신호는 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 또는 우측 정상 채널을 포함하는, 방법.

**청구항 19**

제15항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 디커플링 툴을 적용하는 단계를 포함하는, 방법.

**청구항 20**

제19항에 있어서, 상기 디커플링 툴은:

명시된 주파수 아래의 독립적으로 디코딩된 신호를 추출하는 단계;

상기 명시된 주파수 위의 복합 신호를 추출하는 단계;

상기 복합 신호 그리고 상기 복합 신호의 에너지 및 개개의 신호의 에너지에 관한 스케일 팩터로부터 상기 명시된 주파수 위의 개개의 디커플링된 신호를 결정하는 단계; 및

상기 공동으로 디코딩된 신호를 획득하기 위해 각각의 독립적으로 디코딩된 신호를 개개의 디커플링된 신호와 결합하는 단계를 포함하는, 방법.

**청구항 21**

제15항에 있어서, 상기 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 둘 초과의 오디오 신호를 추출하기 위해 공동-디코딩 툴을 적용하는 단계를 포함하는, 방법.

**청구항 22**

제11항에 있어서, 상기 재생 디바이스와 연관된 상기 하나 이상의 오디오 신호를 디코딩하는 단계는 상기 오디오 신호가 코딩된 것과 동일한 도메인(domain)의 상기 오디오 신호에 대역폭 확장을 적용하는 단계를 포함하는, 방법.

**청구항 23**

제22항에 있어서, 상기 도메인은 변형 이산 코사인 변환(modified discrete cosine transform, MDCT) 도메인인, 방법.

**청구항 24**

제22항 또는 제23항에 있어서, 상기 대역폭 확장은 적응적 잡음 추가를 포함하는, 방법.

**청구항 25**

제1항 내지 제24항 중 어느 한 항의 방법을 수행하도록 구성된 장치.

**청구항 26**

실행될 때, 하나 이상의 디바이스가 제1항 내지 제24항 중 어느 한 항의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

## *발명의 설명*

### *기 술 분 야*

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2022년 10월 5일에 출원된 미국 가출원 제63/378,500호, 및 2023년 8월 24일에 출원된 미국 가출원 제63/534,458호의 우선권의 이익을 주장하고, 이들 각각은 그 전체가 본원에 참조에 의해 통합된다.

[0003] 기술분야

[0004] 본 개시는 일반적으로 오디오 신호 프로세싱에 관한 것으로, 보다 구체적으로는 디바이스들 간의 몰입형(immersive) 오디오 프로그램의 오디오 신호의 낮은 레이턴시(low latency) 교환을 위한 오디오 소스 코딩 및 디코딩에 관한 것이다.

## 배 경 기 술

[0005] 오디오의 스트리밍은 오늘날의 사회에서 일반적이다. 오디오 스트리밍은 사용자의 품질에 대한 상승하는 기대와 함께 더욱 까다로워지고 있지만, 사용자의 설정은 스피커의 수뿐만 아니라 스피커의 유형에 따라 복잡해지고 있다. 스트리밍은 일반적으로 일부 부분에서 적어도 무선 링크를 통해 행해지고, 이는 그 후 양호한 품질을 갖기 위해 무선 링크에 일부 요건을 두며 아마도 많은 이들이 경험한 바와 같이, 이는 항상 그런 것은 아니다.

[0006] 따라서, 특정한 포맷이 클라우드(cloud)/서버(server)로부터 스트리밍되고 후속하여 무선(또는, 일부 경우에 유선) 링크를 통한 분배를 위해 보다 적합한 낮은 레이턴시 포맷으로 트랜스코딩된 디바이스 상에서 스트리밍되는 사용-사례에 대한 교환 포맷을 정의할 필요가 있다. 예시적인 사용-사례는 가정용(in-home) 연결뿐만 아니라, 폰-대-자동차 연결이지만, 포맷은 단일 디바이스로부터 하나 이상의 연결된 디바이스로의 오디오 신호의 낮은 레이턴시 분배가 요구되는 임의의 시나리오에서 유익할 수 있다.

[0007] 무선으로 송신되고 스트리밍되는 오디오 정보 외에, 스트림에 통합되는 다른 유형의 정보가 또한 있을 수 있다. 이러한 다른 유형의 정보는 그 후 또한 무선 링크의 품질에 의해 영향을 받을 것이고 오디오에 대해서와 같은 유사한 단점을 가질 수 있다.

[0008] 따라서, 상이한 유형의 정보 또는 신호와 결합된 상이한 유형의 스트리밍된 오디오에 대한 무선 스트리밍과 연관된 문제를 극복하는 것이 유리할 것이다.

## 발명의 내용

[0009] 본 개시의 목적은 상이한 유형의 정보와 결합된 오디오의 무선 스트리밍에 의해 적어도 부분적으로 위의 문제를 극복하는 것이다.

[0010] 본 개시의 제1 양태에 따르면, 복수의 오디오 신호를 포함하는 오디오 프로그램의 인코딩된 비트스트림의 프레임을 생성하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은, 복수의 오디오 신호 각각에 대해, 개개의 오디오 신호가 연관된 재생(playback) 디바이스를 나타내는 정보를 수신하는 단계, 하나 이상의 인코딩된 오디오 신호를 획득하기 위해 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계, 개개의 재생 디바이스와 연관된 하나 이상의 인코딩된 오디오 신호를 프레임의 제1 독립적인 블록에 결합하는 단계, 복수의 오디오 신호 중 하나 이상의 다른 오디오 신호를 하나 이상의 추가적인 독립적인 블록으로 인코딩하는 단계, 및 제1 독립적인 블록 및 하나 이상의 추가적인 독립적인 블록을 인코딩된 비트스트림의 프레임에 결합하는 단계를 포함한다.

[0011] 본 개시의 제2 양태에 따르면, 인코딩된 비트스트림의 프레임으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은, 인코딩된 비트스트림으로부터, 재생 디바이스와 연관된 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 독립적인 블록을 식별하는 단계, 인코딩된 비트스트림으로부터, 인코딩된 데이터의 식별된 독립적인 블록을 추출하는 단계, 하나 이상의 디코딩된 오디오 신호를 획득하기 위해 인코딩된 데이터의 독립적인 블록으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하는 단계, 인코딩된 비트스트림으로부터, 하나 이상의 추가적인 오디오 신호에 대응하는 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 식별하는 단계, 및 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 디코딩하거나 스킵하는 단계를 포함한다.

[0012] 본 개시의 제3 양태에 따르면, 제1 양태 및/또는 제2 양태의 방법을 수행하도록 구성된 장치이다.

[0013] 본 개시의 제4 양태에 따르면, 실행될 때, 하나 이상의 디바이스가 제1 및/또는 제2 양태의 방법을 수행하는 것

을 야기하는 명령의 시퀀스를 포함하는 비일시적 컴퓨터 판독가능 저장 매체이다.

[0014] 본 개시의 추가적인 예는 종속 청구항에 정의된다.

[0015] 일부 예에서, 둘 이상의 오디오 신호는 재생 디바이스와 연관되고, 둘 이상의 오디오 신호 각각은 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된(bandlimited) 신호이고, 대역제한된 신호 각각에 대해 상이한 인코딩 기술이 사용된다. 일부 예에서, 대역제한된 신호 각각에 대해 상이한 심리음향 모델(psychoacoustic model) 및/또는 상이한 비트 할당 기술이 사용된다. 예에서, 인코딩된 신호의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델(buffer fullness model)에 의해 제약된다.

[0016] 일부 예에서, 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계는 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 프레임의 제1 독립적인 블록으로 공동으로 인코딩하는 단계를 포함한다. 예에서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 이상의 오디오 신호에 걸쳐 하나 이상의 스케일 팩터(scale factor)를 공유하는 단계를 포함한다. 일부 예에서, 둘 이상의 오디오 신호는 공간적으로 관련된다. 예에서, 둘 이상의 공간적으로 관련된 오디오 신호는 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 또는 우측 정상 채널을 포함한다.

[0017] 일부 예에서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 커플링 툴을 적용하는 단계를 포함한다. 일부 예에서, 커플링 툴은 둘 이상의 오디오 신호를 명시된 주파수 위의 복합 신호로 결합하는 단계; 및 둘 이상의 오디오 신호 각각에 대해, 복합 신호의 에너지 및 각각의 개개의 신호의 에너지에 관한 스케일 팩터를 결정하는 단계를 포함한다. 일부 예에서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 초과의 신호에 공동-코딩 툴(joint-coding tool)을 적용하는 단계를 포함한다.

[0018] 일부 예에서, 둘 이상의 오디오 신호는 재생 디바이스와 연관되고, 둘 이상의 오디오 신호 각각은 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된 신호이고, 상이한 디코딩 기술이 둘 이상의 오디오 신호를 디코딩하는 데 사용된다. 예에서, 상이한 심리음향 모델 및/또는 상이한 비트 할당 기술이 대역제한된 신호 각각을 인코딩하기 위해 사용된다.

[0019] 일부 예에 대해, 인코딩된 비트스트림의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델에 의해 제약된다. 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하는 단계는 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 인코딩된 데이터의 독립적인 블록으로부터 공동으로 디코딩하는 단계를 포함한다.

[0020] 예에서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 디커플링 툴을 적용하는 단계를 포함한다. 예에서, 디커플링 툴은, 명시된 주파수 아래의 독립적으로 디코딩된 신호를 추출하는 단계, 명시된 주파수 위의 복합 신호를 추출하는 단계, 복합 신호 그리고 복합 신호의 에너지 및 개개의 신호의 에너지에 관한 스케일 팩터로부터 명시된 주파수 위의 개개의 디커플링된 신호를 결정하는 단계, 및 공동으로 디코딩된 신호를 획득하기 위해 각각의 독립적으로 디코딩된 신호를 개개의 디커플링된 신호와 결합하는 단계를 포함한다.

[0021] 본 개시에서, 프레임은 모든 신호 전체의 시간 슬라이스(time slice)를 나타낸다. 블록 스트림은 세션의 지속기간 동안 신호의 집합을 나타낸다. 블록은 블록 스트림의 하나의 프레임을 나타낸다. 주어진 샘플링 주파수를 갖는 디지털 오디오의 경우, 프레임 크기는 임의의 오디오 신호에 대한 프레임의 오디오 샘플의 수와 동일하다. 프레임 크기는 보통은 세션의 지속기간 동안 일정하게 유지된다.

[0022] 웨이크 워드(wake word)는 하나의 단어, 또는 고정된 순서로 둘 이상의 단어를 포함하는 구를 포함할 수 있다.

[0023] 청구범위를 포함하여 본 개시 전체에 걸쳐, "시스템"이라는 표현은 디바이스, 시스템 또는 서브시스템을 나타내는 넓은 의미로 사용된다. 예를 들어, 디코더를 구현하는 서브시스템은 디코더 시스템으로 지칭될 수 있으며, 이러한 서브시스템을 포함하는 시스템(예를 들어, 다수의 입력에 응답하여 X 출력 신호를 생성하는 시스템, 여기에서 서브시스템은 M개의 입력을 생성하고 나머지 X-M개의 입력은 외부 소스로부터 수신됨)은 또한 디코더 시스템으로 지칭될 수 있다.

## *도면의 간단한 설명*

[0024] 본 개시의 예가 첨부된 도면을 참조하여 상세히 설명될 것이다. 이하의 도면에서, 유사한 참조 번호는 유사한 요소를 지칭하기 위해 사용된다. 이하의 도면은 다양한 예를 묘사하지만, 하나 이상의 구현은 도면에 묘사된 예로 제한되지 않는다.

도 1은 가정용 연결 낮은 레이턴시 트랜스코딩의 예를 예시한다.

도 2는 자동차 연결 오디오 스트리밍의 예를 예시한다.

도 3은 무선 디바이스의 사용에 의한 증강 TV 오디오 스트리밍의 예를 예시한다.

도 4는 간단한 무선 스피커에 의한 오디오 스트리밍의 예를 예시한다.

도 5는 비트스트림 요소로부터 디바이스로의 메타데이터 맵핑(mapping)의 예를 예시한다.

도 6은 오디오 스트리밍의 간단한 가요성 렌더링(flexible rendering)을 어떻게 배치하는지의 예를 예시한다.

도 7은 비트스트림 요소에 대한 가요성 렌더링 데이터의 맵핑의 예를 예시한다.

도 8은 다수의 디바이스로 오디오를 재생하고 커맨드를 청취할 때 에코-참조(echo-reference)의 시그널링(signaling)의 예를 예시한다.

도 9는 프레임, 블록, 및 패킷이 어떻게 서로 관련되는지의 예를 예시한다.

도 10은 현재 MPEG-4 구조의 추가적인 코덱 정보의 예를 예시한다.

도 11은 현재 MPEG-4 구조의 또 다른 추가적인 코덱 정보의 예를 예시한다.

도 12는 청취 능력 및 음성 인식을 통합하는 예를 예시한다.

도 13은 다수의 블록을 포함하는 프레임의 예를 예시한다.

도 14는 다수의 블록을 포함하는 비트스트림의 예를 예시한다.

도 15는 상이한 우선순위를 갖는 블록의 예를 예시한다.

도 16은 상이한 우선순위를 갖는 다수의 블록을 포함하는 비트스트림의 예를 예시한다.

도 17은 상이한 우선순위를 갖는 프레임의 예를 예시한다.

도 18은 상이한 우선순위를 갖는 프레임을 포함하는 비트스트림의 예를 예시한다.

## *발명을 실시하기 위한 구체적인 내용*

[0025] 본 발명의 원리는 이제 도면에 예시된 다양한 예를 참조하여 설명될 것이다. 이러한 예의 묘사는 단지 통상의 기술자에게 본 발명을 더 잘 이해하고 추가로 구현할 수 있게 하기 위한 것임이 이해되어야 하며; 이는 본 발명의 범주를 어떠한 방식으로도 제한하는 것으로 의도되지 않는다.

[0026] 도 1에서, 몰입형 오디오 스트림은 클라우드 또는 서버(10)로부터 스트리밍되고, TV 또는 허브 디바이스(20) 상에서 디코딩된다. 몰입형 오디오 스트림은, 예를 들어, Dolby Digital Plus, AC-4 등을 포함하는 임의의 기존 포맷으로 코딩될 수 있다. 출력은, 바람직하게는 로컬 무선 연결, 예를 들어, WiFi 소프트 액세스 포인트, 또는 블루투스 연결을 통해 연결된, 연결된 디바이스(30)로의 추가 전송을 위해 낮은 레이턴시 교환 포맷으로 후속적으로 트랜스코딩된다. 낮은 레이턴시는 보통은, 예를 들어 프레임 크기, 샘플링 레이트, 하드웨어 및/또는 소프트웨어 계산 자원 등과 같은 다양한 팩터에 의존하지만, 낮은 레이턴시는 일반적으로 40ms, 20ms 또는 10ms 미만일 것이다.

[0027] 또 다른 통상적인 사용-사례에서, 폰은 클라우드 또는 서버로부터 몰입형 오디오 스트림을 가져와서 교환 포맷으로 트랜스코딩하고 후속하여 연결된 자동차로 전송한다. 도 2의 예시된 예에서, 모바일 디바이스(예를 들어, 폰 또는 태블릿)(20)는 서버(10)에 연결되어 몰입형 오디오 스트림을 수신하고, 낮은 레이턴시 트랜스코드를 낮은 레이턴시 교환 포맷에 수행하고, 트랜스코딩된 신호를 몰입형 오디오 재생을 지원하는 자동차(30)에 전송한다. 몰입형 오디오 스트림의 예는 Dolby Atmos 포맷의 오디오를 포함하는 스트림이고, 몰입형 오디오 재생을 지원하는 자동차(30)의 예는 Dolby Atmos 몰입형 포맷을 재생하도록 구성된 자동차이다.

[0028] 일반적으로, 교환 포맷은 바람직하게는, 낮은 레이턴시, 낮은 인코딩 및 디코딩 복잡도, 높은 품질로 스케일링하는 능력, 및 합리적인 코딩 효율을 갖는다. 포맷은 또한 바람직하게는, 구성 가능한 레이턴시를 지원하여서,

다양한 연결 조건 하에서 동작할 수 있도록 레이턴시가 효율, 및 또한 에러 복원력(error resilience)과 거래될 수 있게 된다.

[0029] 청취 능력이 있거나 없는 무선 스피커가 있는 허브 또는 증강 디스플레이

[0030] 도 3에 도시된 것은 무선 스피커(30) 또는 디스플레이(예를 들어, 텔레비전 또는 TV)(20)의 세트를 구동하는 허브(20)이고, 가능한 경우 내장 스피커(30)가 여러 개의 무선 스피커(30)에 의해 증강된다. 증강은, 디스플레이(20)가 스피커(30)를 포함하는 예에서, 디스플레이(20)가 마찬가지로 오디오 재현의 부분이라는 것을 제안한다. 무선 스피커/디바이스(30)는 동일한 완전한 신호를 수신할 수 있거나(브로드캐스트 모드(Broadcast Mode), 도 3에서 좌측에 예시됨) 또는 특정 디바이스에 대해 맞춤화된 개별 스트림(유니캐스트-멀티포인트 모드(Unicast-multipoint Mode), 도 3에서 우측에 예시됨)을 수신할 수 있다.

[0031] 각각의 스피커는 동일한 채널의 상이한 주파수 범위를 커버하거나, 표준(canonical) 맵핑에서 상이한 채널에 대응하는 다수의 드라이버를 포함할 수 있다. 예를 들어, 스피커는 두 개의 드라이버를 가질 수 있고, 그 중 하나는 상승된 스피커를 에뮬레이팅하기 위해 높이 채널에 대응하는 신호를 출력하는 상방 발사 드라이버일 수 있다. 무선 디바이스는 청취-능력을 가질 수 있고(예를 들어, "스마트 스피커"), 따라서, 스피커가 하나 이상의 에코-참조(echo-reference)를 수신할 것을 요구할 수 있는 에코-관리를 요구할 수도 있다. 에코-참조는 (예를 들어, 동일한 스피커/디바이스에 대해) 로컬일 수 있거나, 대신에 근방의 다른 스피커/디바이스로부터의 관련 신호를 나타낼 수 있다.

[0032] 스피커는 임의의 위치에 배치될 수 있고, 이 경우, 렌더링이 스피커의 실제 포지션을 고려하는(예를 들어, 라우드스피커(loudspeaker)는 고정되고 미리 정의된 위치에 위치되는 것으로 가정되는 표준 가정과는 대조적임) 소위 "가요성 렌더링"이 수행될 수 있다. 가요성 렌더링은 허브 또는 TV에서 일어날 수 있고, 후속적으로 렌더링된 신호 각각이 각각의 디바이스에 전송되고, 각각의 개개의 디바이스가 적절한 신호를 추출하고 출력하는 브로드캐스트 모드로, 또는 개별 디바이스에 대한 개별 스트림으로서 렌더링된 신호가 스피커/디바이스에 송신된다. 대안적으로, 가요성 렌더링은 각각의 디바이스 상에서 국부적으로 일어날 수 있고, 이에 의해 각각의 디바이스는 완전한 몰입형 프로그램의 표현, 예를 들어, 7.1.4 채널 기반 몰입형 표현을 수신하고, 그 표현으로부터 개개의 디바이스에 적합한 출력 신호를 렌더링한다.

[0033] 무선 디바이스는 디바이스에 의해 제공된 소프트 액세스 포인트를 통해 또는 가정 내의 로컬 액세스 포인트를 통해 허브 또는 TV에 연결될 수 있다. 이는 비트레이트 및 레이턴시에 상이한 요건을 부과할 수 있다.

[0034] 자동차 애플리케이션의 모바일 디바이스 프로젝션

[0035] 또 다른 사용-사례에서, 모바일 디바이스(예를 들어, 폰 또는 태블릿)(20)는 클라우드 또는 서버(10)로부터 몰입형 오디오 스트림을 가져와서, 몰입형 오디오 스트림을 교환 포맷으로 트랜스코딩하고, 후속하여 트랜스코딩된 스트림을 연결된 자동차(30)로 전송한다. 도 2의 예에서, Dolby Atmos-인에이블드(enabled) 폰(20)은 낮은 레이턴시 트랜스코딩 및 Dolby Atmos-인에이블드 자동차(3)로의 전송을 위해 서버(10)에 연결되고 Dolby Atmos 스트림을 수신한다. 이러한 사용-사례에서, 거실 사용-사례에서보다 더 높은 비트레이트가 이용 가능할 수 있고, 또한 무선 채널의 상이한 특징이 존재할 수 있다. 자동차 사용-사례에서 더 높은 비트레이트에 대한 가능한 이유는 이웃, 다른 방 등과 같은 모든 무선 디바이스가 무선 환경에 잡음을 추가하는 거실에 비교하여, 자동차가 일종의 패러데이 케이지(faraday cage)의 역할을 하고 외부 무선 방해로부터 내부 자동차 환경을 차폐하기 때문에 덜 잡음이 있는 무선 환경 때문이다. 동시에, 일반적으로, 거실의 무선 디바이스를 포함하는, 가정에 있는 모든 무선 디바이스가 이용 가능한 대역폭에 대해 경쟁하는 거실 사용 사례와 비교하여, 일반적으로는 이용 가능한 무선 대역폭에 대해 경쟁하는 매우 적은 무선 디바이스가 자동차에 존재한다.

[0036] 이러한 사용-사례의 경우, 모바일 디바이스에 의해 트랜스코딩되고 자동차로 전송될 신호는 채널 기반 몰입형 표현, 객체 기반 표현, 장면 기반(scene-based) 표현(예를 들어, 및 Ambisonics 표현), 또는 심지어 상이한 표현의 조합일 수 있다는 것이 상상된다. 이 예의 경우, 상이한 렌더링 아키텍쳐 및 브로드캐스트 대 멀티포인트는 관련성이 없을 수 있는데, 이는 일반적으로 완전한 프리젠테이션이 모바일 디바이스로부터 단일 엔드-포인트(예를 들어, 자동차)로 전송될 것이기 때문이다.

[0037] 교환 포맷의 설명

[0038] 몰입형 교환 포맷은 지각적으로(perceptually) 동기 부여된 양자화 및 코딩에 의해 변형 이산 코사인 변환(modified discrete cosine transform, MDCT) 상에 구축된다. 이는 구성 가능한 레이턴시, 예를 들어, 주어진 샘플링 레이트에서 상이한 변환 크기에 대한 지원을 갖는다. 예시적인 프레임 크기는 48kHz 및 44.1kHz의 샘플

링 레이트에서 128, 256, 512, 1024 및 120, 240, 480, 960 및 192, 384, 768 샘플이다.

[0039] 포맷은 몰입형 채널 구성(예를 들어, 5.1.2, 5.1.4, 7.1.2, 7.1.4, 9.1.6, 및 22.2, 또는 ISO/IEC 23091-3:2018, 표-2에 명시된 바와 같은 고유한 채널로 이루어진 임의의 다른 채널 구성을 포함하지만 이에 제한되지 않음)을 포함하는 모노, 스테레오, 5.1 및 다른 채널 구성을 지원할 수 있다. 포맷은 또한 Ambisonics(예를 들어, 1차 이상)와 같은 객체 기반 오디오 및 장면 기반 표현을 지원할 수 있다. 포맷은 (예를 들어, MPEG-4 오디오 표준으로 또한 지칭될 수 있는 ISO/IEC 14496-3, MPEG-4 시스템 표준으로 또한 지칭될 수 있는 ISO 14496-1, ISO 베이스 미디어 파일 포맷 표준으로 또한 지칭될 수 있는 ISO 14496-12 및/또는 MP4 파일 포맷 표준으로 또한 지칭될 수 있는 ISO 14496-14에 설명된 것들과 같은) 기존 포맷과의 통합에 적합한 시그널링 체계를 또한 사용할 수 있다.

[0040] 또한, 시스템은 신택스 요소의 부분을 스킵하고 주어진 스피커에 대한 관련 부분만을 디코딩하는 능력에 대한 지원, 지연 정렬, 레벨 조정, 및 동등화(equalization)와 같은 메타데이터 제어된 가요성-렌더링 양태에 대한 지원, 청취 능력, 및 에코-참조의 시그널링에 의한 에코-관리에 대한 연관된 지원을 갖는 (스마트 스피커의 드라이버/라우드스피커 각각을 독립적으로 피드하는 것을 허용하는 신호의 세트가 송신되는 시나리오를 포함하는) 스마트 스피커의 사용에 대한 지원, 낮은 중첩 윈도우뿐만 아니라 50% 중첩 윈도우를 갖는 MDCT 도메인에서의 양자화 및 코딩에 대한 지원, 및/또는 양자화 잡음의 시간적 성형, 예를 들어, 시간 잡음 성형(Temporal Noise Shaping, TNS)을 하기 위해 MDCT 도메인의 주파수 축을 따른 필터링에 대한 지원을 가질 수 있다. 따라서, 다양한 예에서, 중첩 윈도우는 대칭 또는 비대칭이다.

[0041] 개시된 교환 포맷은 일부 예에서 개선된 공동 채널 코딩을 제공하여, 가요성 렌더링 사용-사례에서의 신호들 사이의 증가된 상관성, 채널 요소에 걸친 공유된 스케일 팩터에 의한 개선된 코딩 효율, MDCT 도메인에서 고주파수 재구성 및 잡음 추가 기술의 포함, 보다 양호한 효율을 허용하기 위한 레거시 코딩 구조에 대한 분류된 개선 및 가까운 사용-사례에 대한 적합성의 이점을 취할 수 있다.

[0042] 브로드캐스트 모드에서 개별 재생 디바이스를 제어하는 일부 예에서, 스킵 가능한 블록 및 메타데이터가 사용될 수 있다. 브로드캐스트 모드 설정에서, 각각의 무선 디바이스는 특정 디바이스에 대응하는 특정 드라이버/채널에 대한 완전한 오디오의 관련 부분을 재생하는 것이 필요할 것이다. 이는 디바이스가 스트림의 어떤 부분이 그 디바이스와 관련이 있는지 알고, 브로드캐스트되고 있는 완전한 오디오 스트림으로부터 그 부분을 추출할 필요가 있음을 의미한다. 낮은 복잡도 디코딩 동작을 가능하게 하기 위해, 특정 디바이스에 대한 디코더가 디코딩에 관련되지 않은 요소를 지나쳐서 디바이스 및 디바이스 상의 주어진 드라이버에 관련된 요소로 효율적으로 스킵할 수 있는 방식으로 스트림이 구성되는 것이 바람직하다.

[0043] 하지만, 상이한 디바이스(예를 들어, 지나치게 단순화한 시나리오에서, 스테레오 프리젠테이션의 좌측 및 우측 채널)로 예정된 신호들 사이의 공동 코딩을 여전히 허용하는 것이 (압축 관점으로부터) 유익함이 존재할 수 있다는 것에 유의해야 한다.

[0044] 그러므로, 포맷은: 특정 디바이스에 대해 관련된 부분만의 효율적인 디코딩을 가능하게 하기 위해 비트스트림의 "스킵 가능한 블록"을 가능하게 할 수 있고, 상이한 스피커/디바이스에 대응하는 신호들 간에 공동 코딩 기술을 적용하는 능력을 유지하면서, 특정 디바이스에 대한 하나 이상의 스킵 가능한 블록의 가요성 맵핑을 가능하게 하는 메타데이터를 포함할 수 있다.

[0045] 도 4에서, 임의의 셋업의 예가 도시된다. 세 개의 연결된 무선 디바이스들(31, 32, 33)이 있으며, 이들 중 두 개는 단일 채널 스피커들(32, 33)(오디오의 하나의 채널을 의미함)이고, 하나의 스피커(31)는 3개의 상이한 드라이버를 갖고 이에 의해 세 개의 상이한 신호에서 동작하는 더 진보된 스피커(31)이다. 제1 스피커(31)는 이 예에서 세 개의 개별 신호에서 동작하는 반면, 제2 및 제3 스피커(32, 33)는 스테레오 표현(예를 들어, 좌측 및 우측)에서 동작한다. 그리하여, 스테레오 표현의 신호의 공동 코딩을 하는 것이 유익할 수 있다.

[0046] 위의 시나리오가 주어지면, 포맷은 스킵 가능한 블록을 포함하는 비트스트림을 명시하여서, 제1 디바이스가 그 스피커에 대한 신호를 디코딩하기 위한 스트림의 관련 부분만을 추출할 수 있는 반면, 각각의 스피커가 단일 신호를 출력하기 위해 두 개의 신호를 디코딩할 필요가 있으면서 공동 코딩이 수행될 수 있는 장점이 있는 신호를 구성함으로써 스테레오 쌍에 대해 디코더 복잡도 대 효율이 절충(traded off)될 수 있다.

[0047] 포맷은 복수의 스킵 가능한 블록 중 특정 블록을 하나 이상의 디바이스에 맵핑하는 일반 및 가요성 표현을 가능하게 하기 위해 메타데이터 포맷을 명시한다. 이는 도 5에 예시되며, 여기서 맵핑이 각각의 디바이스(31, 32, 33)를 하나 이상의 비트스트림 요소에 연관시키는 매트릭스로서 표현될 수 있어서, 주어진 디바이스에 대한 디

코더가 어떤 비트스트림 요소를 출력 및 디코딩할지를 알게 될 것이다.

[0048] 예를 들어, 도 5의 예에서, 제1 블록 또는 스킵 블록(Blk1)은 세 개의 단일 채널 요소(디바이스 1(31)의 각각의 드라이버에 대해 하나씩)를 포함하는 한편, 제2 블록 또는 스킵 블록(Blk2)은 디바이스 2(32) 및 디바이스 3(33)에 의해 출력될 신호의 공동으로 코딩된 버전을 포함할 수 있는 채널 쌍 요소를 포함한다. 디바이스 1(31)은 맵핑 메타데이터를 추출하고, 맵핑 메타데이터가 요구하는 신호가 스킵 블록 1(Blk1)에 있다고 결정한다. 따라서, 이는 스킵 블록 1(Blk1)을 추출하고, 내부의 세 개의 단일 채널 요소를 디코딩하고, 이를 드라이버 1, 드라이버 2a, 및 드라이버 2b에 각각 제공한다.

[0049] 또한, 디바이스 1(31)은 스킵 블록 2(Blk2)를 무시한다. 유사하게, 디바이스 2(32)는 맵핑 메타데이터를 추출하고, 맵핑 메타데이터가 요구하는 신호가 스킵 블록 2(Blk2)에 있다고 결정한다. 따라서, 디바이스 2(32)는 스킵 블록 1(Blk1)을 스킵하고, 스킵 블록 2(Blk2)을 추출한다. 디바이스 2(32)는 채널 쌍 요소를 디코딩하고 CPE의 좌측 채널 출력을 그의 드라이버에 제공한다. 유사하게, 디바이스 3(33)은 맵핑 메타데이터를 추출하고, 맵핑 메타데이터가 요구하는 신호가 스킵 블록 2(Blk2)에 있다고 결정하고, 따라서 디바이스 3(33)은 스킵 블록 1(Blk1)을 스킵하고, 또한 스킵 블록 2(Blk2)를 추출한다. 디바이스 3(33)은 채널 쌍 요소를 디코딩하고 CPE의 우측 채널 출력을 그의 드라이버에 제공한다.

[0050] 일부 예에서, 디바이스 2(32)는 스킵 블록 2(Blk2)로부터의 신호의 서브세트만을 요구한다고 결정할 수 있다. 이러한 예에서, 가능할 때, 디바이스 2(32)는 스킵 블록 2(Blk2)에서 신호를 완전히 디코딩하는데 요구되는 동작의 서브세트만을 수행할 수 있다. 구체적으로, 도 5의 예에서, 디바이스 2(32)는 CPE의 좌측 채널을 추출하기 위해 요구되는 프로세싱 동작만을 수행할 수 있고, 따라서 계산 복잡도의 감소를 가능하게 한다. 유사하게, 디바이스 3(33)은 CPE의 우측 채널을 추출하기 위해 요구되는 프로세싱 동작만을 수행할 수 있다.

[0051] 예를 들어, 도 5의 경우, CPE가 공동-채널 코딩을 사용하여 코딩되는 시간, 및 CPE가 독립적인 채널 코딩을 사용하여 코딩되는 다른 시간이 있을 수 있다. CPE가 독립적인 채널 코딩을 사용하여 코딩될 때의 시간 동안, 디바이스 2(32)는 CPE의 제1(예를 들어, 좌측) 채널만을 추출할 수 있는 반면, 디바이스 3(33)은 CPE의 제2(예를 들어, 우측) 채널만을 추출할 수 있다.

[0052] 다른 예에서, CPE의 채널은 공동-채널 코딩을 사용하여 코딩될 수 있으며, 이 경우, 디바이스 2(32) 및 디바이스 3(33)은 CPE의 두 개의 중간 채널을 추출해야만 한다. 하지만, 디바이스 2(32)는 중간 채널로부터 좌측 채널을 추출하기 위해 요구되는 그 동작만을 수행함으로써 감소된 계산 복잡도로 동작하는 것이 여전히 가능할 수 있다. 유사하게, 디바이스 3(33)은 중간 디코딩된 채널들로부터 우측 채널을 추출하기 위해 요구되는 그 동작만을 수행함으로써 감소된 계산 복잡도로 동작하는 것이 여전히 가능할 수 있다.

[0053] 공동-채널 코딩의 세부 사항에 따라, 다른 최적화가 가능할 수 있다. 상이한 디바이스/디코더 각각의 아이덴티티는 시스템 초기화 또는 셋업 페이즈 동안 정의될 수 있다. 이러한 셋업은 공통적이고 일반적으로는 방의 음향, 스위트 스팟(sweet spot)에 대한 스피커 거리 등을 측정하는 것을 수반한다.

[0054] 가요성 렌더링(프리-인코딩 및 포스트 디코딩)의 분배 및 지연 적용 등 포스트 디코더의 적용

[0055] 특정 디바이스/스피커에 관련 신호를 송신하기 전에 허브/TV에서 가요성 렌더링이 적용되는 사용-사례에 있어서, 렌더링은, 코딩 관점으로부터, 예를 들어, 해당 신호의 공동 코딩을 고려할 때, 코딩하기가 더 어려운 신호를 생성할 수 있다. 하나의 이유는, 가요성 렌더링이 (예를 들어, 다른 스피커 및 청취자에 대한 스피커의 배치에 따라) 상이한 디바이스에 대해 상이한 지연, 동등화, 및/또는 이득 조정을 적용할 수 있기 때문이다. 정보, 예를 들어, 초기 셋업에서의 이득 및 지연과 같은 정보를 미리 설정하고, 유연하게 동등화만을 렌더링하는 것이 또한 가능할 것이다. 미리 설정된 정보 및 가요성 렌더링 정보의 다른 변형이 또한 다른 예에서 가능할 수 있다. 본 문서에서, "이득"이라는 용어는, 단지 특정한 레벨 조정(예를 들어, 증폭)으로 제한되기 보다는, 임의의 레벨 조정(예를 들어, 감쇠, 증폭 또는 관통(pass-through))을 의미하는 것으로 해석되어야 한다는 것에 유의한다.

[0056] 도 6의 예에서, 우측 채널(33) 및 좌측 채널(32) 스피커에는 청취자에 대한 상이한 배치를 반영하는 상이한 레이턴시가 주어진다(예를 들어, 스피커들(32, 33)이 청취자에 대해 등거리가 아닐 수 있기 때문에, 상이한 스피커들(32, 33)로부터 일관된(coherent) 사운드가 동시에 청취자에 도달하도록 상이한 레이턴시가 스피커들(32, 33)로부터 출력된 신호에 적용될 수 있음). 상이한 스피커들(31, 32, 33)을 통한 재생을 위해 의도된 일관된 신호에 대한 이러한 상이한 레이턴시의 도입은 이러한 신호의 공동 코딩을 어렵게 한다.

[0057] 이러한 어려움을 처리하기 위해, 가요성 렌더링 프로세스의 양태는 신호의 디코딩 후에 엔드포인트 디바이스에

서 파라미터화되고 적용될 수 있다.

[0058] 도 7의 예에서, 각각의 디바이스에 대한 지연 및 이득 값은 파라미터화되고 개개의 스피커들(31, 32, 33)로 송신되는 인코딩된 신호에 포함된다. 개개의 신호는 개개의 디바이스들(31, 32, 33)에 의해 디코딩될 수 있고, 이는 그 후 파라미터화된 이득 및 지연 값을 개개의 디코딩된 신호에 도입할 수 있다.

[0059] 상이한 디바이스들(31, 32, 33)에 대한 코딩된 신호가 분리 가능한 블록(예를 들어, 스킵 가능한 블록)에서 송신되는, 도 7에 도시된 바와 같은 예에서, 파라미터(예를 들어, 지연 및 이득)는 또한 분리 가능한 블록에서 송신될 수 있어서, 디바이스(31, 32, 33)가 그 디바이스(31, 32, 33)에 대해 요구되는 파라미터의 서브세트만을 추출하고 그 디바이스(31, 32, 33)에 대해 요구되지 않는 이러한 파라미터를 무시(및 스킵)할 수 있다. 이러한 경우, 어떤 블록에 어떤 파라미터가 포함되는지를 표시하는 맵핑 메타데이터가 각각의 디바이스에 제공될 수 있다.

[0060] 도 7은 지연 및 이득 파라미터만을 표시하지만, 동등화 파라미터와 같은 다른 파라미터가 또한 포함될 수 있다는 것에 또한 유의한다. 동등화 파라미터는, 예를 들어, 상이한 주파수 영역에 적용될 복수의 이득, 재생 디바이스(31, 32, 33)에 의해 적용될 미리 결정된 동등화 곡선의 표시, 무한 임펄스 응답(infinite impulse response, IIR) 또는 유한 임펄스 응답(finite impulse response, FIR) 필터 계수의 하나 이상의 세트, 바이쿼드(biquad) 필터 계수의 세트, 파라메트릭 동등화기의 특징을 명시하는 파라미터, 및 통상의 기술자에게 공지된 동등화를 명시하는 다른 파라미터를 포함할 수 있다.

[0061] 또한, 가요성 렌더링 양태의 파라미터화는 정적일 필요가 없고, (예를 들어, 청취자가 오디오 프로그램의 재생 동안 이동하는 경우) 동적일 수 있다. 그리하여, 파라미터가 동적으로 변경되는 것을 허용하는 것이 바람직할 수 있다. 오디오 프로그램 동안 하나 이상의 파라미터가 변경되는 경우에, 디바이스는 매끄러운 전이를 제공하기 위해 이전의 지연 및/또는 이득 파라미터와 업데이트된 지연 및/또는 이득 파라미터 사이에서 보간할 수 있다. 이는 시스템이 청취자의 위치를 동적으로 추적하고, 동적 렌더링을 위해 스위트-스팟을 대응적으로 업데이트하는 상황에서 특히 유용할 수 있다.

[0062] 또한 주지된 바와 같이, 그리고 후속하여 설명될 바와 같이, 가요성 렌더링이 적용될 때, 채널들 간의 증가된 레벨의 상관이 발생할 수 있고, 이는 보다 가요성의 공동 코딩에 의해 이용될 수 있다.

[0063] 에코-참조 코딩 및 시그널링

[0064] 도 8의 좌측에 예시된, 브로드캐스트 모드 또는 도 8의 우측에 예시된 유니캐스트-멀티포인트 모드에서 신호를 수신하여 함께 동작하는 다수의 디바이스/스피커(30)가 있고, 동시에 "청취" 능력을 가능하게 하기 위해 디바이스(30) 상에 마이크로폰(40)이 있는, 도 8에 개략적으로 도시된 바와 같은 사용-사례의 경우, 에코-관리에 대한 필요성이 발생한다.

[0065] 다수의 스피커/디바이스와 관련하여 에코-관리를 수행할 때, 단지 로컬 스피커 디바이스보다 많은 스피커 디바이스로부터의 에코-참조를 사용하는 것이 유익할 수 있다. 예로서, 하나의 디바이스가 다른 디바이스에 가깝게 배치될 수 있고, 그러므로, 가까운 근방의 디바이스로부터의 신호는 활성 마이크로폰에 의해 디바이스의 에코-관리에 영향을 미칠 것이다. 브로드캐스트 모드 사용-사례에서, 각각의 디바이스는 모든 디바이스에 대한 신호를 수신한다. 디바이스가 다른 디바이스에 대한 신호를 가질 때, 그 신호를 에코-참조로서 사용하는 것이 유익할 수 있다. 그렇게 하기 위해, 신호 중 어느 것이 다른 디바이스에 대한 에코-참조로서 사용될 수 있는지를 특정 디바이스에 시그널링할 필요가 있다.

[0066] 예에서, 이는 특정 스피커/디바이스에 의해 (전체 세트로부터) 재생될 채널/신호를 맵핑할 뿐만 아니라, 특정 스피커/디바이스에 대해 (전체 세트로부터) 에코-참조로서 사용될 채널/신호를 맵핑하는 메타데이터를 제공함으로써 행해질 수 있다. 이러한 메타데이터 또는 시그널링은 동적일 수 있으며, 이는 바람직한 에코-참조의 표시가 시간에 따라 변하는 것을 가능하게 한다.

[0067] 각각의 디바이스/스피커가 재생되어야 하는 특정 신호만을 수신하는 사용-사례의 경우, 적절한 에코-참조를 제공하기 위해, 추가적인 신호(예를 들어, 에코-참조 신호)를 각각의 디바이스/스피커에 전송하는 것이 필요할 수 있다. 다시, 그렇게 하기 위해, 각각의 디바이스가 재생을 위한 적절한 신호 및 에코-관리를 위한 적절한 신호를 선택할 수 있도록 디바이스-특정 시그널링을 제공할 필요가 있다.

[0068] 에코 관리를 위한 신호는 에코 관리를 수행하기 위한 디바이스에 의해서만 사용되고, 청취자를 위한 디바이스에 의해서는 재생되지 않기 때문에, 에코-참조 신호는 디바이스에 의한 재생을 위해 의도되는 신호와 상이하게 코

딩되거나 표현될 수 있다. 구체적으로, 성공적인 에코-관리는 청취자를 위한 재생을 위해 일반적으로 사용되는 것보다 더 낮은 레이트로 코딩된 신호로 달성될 수 있기 때문에, 청취자에 대한 재생을 위해 전혀 적합하지 않을 수 있지만, 오디오 신호의 필요한 피처를 캡처하는, 신호의 파라메트릭 표현과 같은 부가적인 압축 툴은 상당히 감소된 전송 비용으로 양호한 에코-관리를 제공할 수 있다.

[0069] 분류된 신택스 요소

[0070] 일부 예에서, 블록은 오디오의 전달을 최적화하기 위해 사용된다. 설명된 포맷에서, 각각의 프레임은 스킵 가능한 블록과 관련하여 전술된 바와 같이, 블록으로 분할될 수 있다. 블록은 블록이 속한 프레임 번호, 동일한 ID를 갖는 상이한 프레임으로부터의 연속적인 블록을 블록 스트림에 연관시키기 위해 사용될 수 있는 블록 ID, 및 재전송을 위한 우선순위에 의해 식별될 수 있다. 위 내용 중 하나의 예는 다수의 프레임들(N-2, N-1 및 N)을 갖는 스트림이고, 프레임 N은 도 13에 예시된 ID1, ID2, ID3으로 식별된 다수의 블록을 포함한다. 비록 비트스트림 포맷이지만 이 스트림의 예는 도 14에 예시된다. 일부 예에서, 몰입형 오디오 프로그램의 오디오 신호에 대해, 블록의 블록 ID는 전체 몰입형 오디오 프로그램의 신호 중 어느 세트가 그 블록에 의해 운반되는지를 표시할 수 있다.

[0071] 설명된 포맷을 위한 사용 사례는 낮은 레이턴시에서 Wifi와 같은 무선 네트워크를 통한 오디오의 신뢰성 있는 전송이다. 예를 들어, Wifi는 패킷 기반 네트워크 프로토콜을 사용한다. 패킷 크기는 일반적으로 제한된다. IP 네트워크에서 통상적인 최대 패킷 크기는 1500 바이트이다. 스트림의 블록 기반 아키텍처는 전송을 위해 패킷을 어셈블링할 때 유연성을 허용한다. 예를 들어, 더 작은 프레임을 갖는 패킷은 다른 프레임으로부터의 재전송된 블록으로 채워질 수 있다. 큰 프레임은 네트워크 프로토콜 계층 상의 패킷들 간의 의존성을 감소시키기 위해 패킷화하기 전에 블록 경계 상에서 분할될 수 있다.

[0072] 도 9는 프레임, 블록, 패킷 간의 관계를 도시한다. 프레임은 시작 시간, 종료 시간, 및 종료 시간과 시작 시간의 차이인 지속기간을 갖는 오디오 신호의 연속적인 세그먼트를 표현하는 오디오 데이터, 바람직하게는 모든 오디오 데이터를 운반한다. 연속 세그먼트는 ISO/IEC 14496-3, 서브파트 4, 섹션 4.5.2.1.1에 따른 시간 기간을 포함할 수 있다. 섹션 4.5.2.1.1은 raw_data_block()의 내용을 설명한다. 프레임은 또한, 예를 들어, 더 낮은 데이터 레이트로 인코딩되는 그 세그먼트의 중복적인 표현을 운반할 수 있다. 인코딩 후, 그 프레임은 블록으로 분할될 수 있다. 블록은 패킷 기반 네트워크를 통한 전송을 위해 패킷으로 결합될 수 있다. 상이한 프레임으로부터의 블록은 단일 패킷으로 결합될 수 있고, 및/또는 비순차적으로 송신될 수 있다.

[0073] 예에서, 블록은 개별 디바이스를 어드레싱(addressing)하기 위해 사용된다. 데이터의 패킷은 개별 디바이스 또는 디바이스의 관련 그룹에 의해 수신된다. 스킵 가능한 블록의 개념은 개별 디바이스 또는 디바이스의 관련 그룹을 어드레싱하기 위해 사용될 수 있다. 패킷을 상이한 디바이스로 전송할 때 네트워크가 브로드캐스트 모드로 동작하더라도, 오디오의 프로세싱(예를 들어, 디코딩, 렌더링 등)은 그 디바이스에 어드레싱된 블록으로 감소될 수 있다. 동일한 패킷 내에서 수신되더라도, 모든 다른 블록은 간단히 스킵될 수 있다. 일부 예에서, 블록은 그의 디코딩 또는 프리젠테이션 시간에 기초하여, 올바른 순서로 가져가질 수 있다. 더 낮은 우선순위를 갖는 재전송된 블록은 더 높은 우선순위를 갖는 동일한 블록이 수신된 경우 제거될 수 있다. 그 후, 블록의 스트림은 디코더로 피드될 수 있다.

[0074] 일부 예에서, 스트림 및 디바이스의 구성은 대역 외(out of band)로 송신된다. 코덱은 오디오 스트림이 비교적 높은 레이트이지만 낮은 레이턴시로 전송되는 연결을 설정하는 것을 허용한다. 이러한 연결의 구성은 이러한 연결의 지속기간에 걸쳐 안정적으로 유지될 수 있다. 그 경우, 오디오 스트림의 구성 부분을 만드는 대신에, 오디오 스트림은 대역 외로 전송될 수 있다. 이러한 대역 외 전송을 위해, 심지어 상이한 네트워크, 또는 네트워크 프로토콜이 사용될 수 있다. 예를 들어, 오디오 스트림은 낮은 레이턴시 전송을 위해 사용자 데이터그램 프로파일(User Datagram Profile, UDP)을 사용할 수 있는 반면, 구성은 구성의 신뢰성 있는 전송을 보장하기 위해 전송 제어 프로토콜(Transmission Control Protocol, TCP)을 사용할 수 있다.

[0075] MPEG-4 오디오 컨텍스트에서 코덱 활성화

[0076] 본 기술의 하나의 특정 애플리케이션은 MPEG-4 오디오 내에서 사용된다. MPEG-4 오디오에서는 코덱 기술에 따라 상이한 AOT(Audio Object Type)가 정의된다. 본 문서에 설명된 포맷에 대해, 새로운 AOT가 정의될 수 있고, 이는 그 포맷에 대한 특정 시그널링 및 데이터를 허용한다. 또한, MPEG-4에서의 디코더의 구성은 DecoderSpecificInfo() 페이로드에서 행해지며, 이는 차례로 AudioSpecificConfig() 페이로드(payload)를 운반한다. 후자에서, 특정 AOT에 대한 특정 정보뿐만 아니라, 샘플링 레이트 및 채널 구성과 같은 특정 포맷에 애그

노스틱한(agnostic) 일반적인 시그널링이 정의된다. 전체 스트림이 단일 디바이스에 의해 디코딩되는 종래의 포맷의 경우, 이는 맞을 수 있다. 하지만, 단일 스트림이 여러 개의 디코더로 전송되는 브로드캐스트 모드에서, 각각의 개개의 디코더가 스트림의 부분만을 디코딩하면서, (디바이스의 출력 기능을 셋업하기 위한 수단으로서) 업-프론트(up-front) 채널 구성 시그널링은 최적이 아닐 수 있다.

[0077] 도 10은 회색의 수정을 갖는, 기존의 MPEG-4 고레벨 구조(검정)를 예시한다. 브로드캐스트 사용-사례를 지원하기 위해, codecSpecificConfig()(여기서 "codec"은 일반적인 플레이스홀더(placeholder) 이름일 수 있음)가 정의되며, 여기서 시그널링은 특정 사용-사례에 대해 재정의되어서, 특정 채널 요소를 특정 디바이스에 맵핑하는 것이 가능할뿐만 아니라, 다른 관련 정적 파라미터를 포함하는 것이 가능하다. "0"의 값을 갖는 MPEG-4 요소 채널 구성은 codecSpecificConfig에 정의된 채널 구성으로 정의된다. 그러므로, 이 값은 codec specific config 내측의 채널 구성의 시그널링의 수정을 가능하게 하기 위해 사용될 수 있다

[0078] 또한, MPEG-4의 취지에서, codecSpecificConfig()가 디코딩 가능하다는 것을 고려하여, 로우(raw) 페이로드가 가까운 특정 디코더에 대해 명시된다. 하지만, 포맷은 동적 메타데이터가 로우 페이로드의 일부임을 보장하고, 모든 로우 페이로드에 대해 길이 정보가 이용 가능한 것을 보장하여서, 디코더는 특정 디바이스에 관련되지 않은 요소를 쉽게 스킵할 수 있다.

[0079] 도 11에서, MPEG-4에서 정의된 예시적인 raw_data_block의 부분이 우측에 주어진다. 로우 데이터 블록은 채널 요소(단일 채널 요소(single channel element, SCE) 또는 채널 쌍 요소(channel pair element, CPE))를 주어진 순서로 포함한다. 하지만, 가까운 출력 디바이스에 관련되지 않을 수 있는 이러한 채널 요소의 부분을 스킵하기를 희망하는 디코더는, 종래의 MPEG-4 오디오 신택스에서, 관련 부분을 추출하는 것이 가능하도록 모든 채널 요소를 파싱(parse)해야 (그리고 어느 정도는 또한 디코딩해야) 할 것이다. 도 11의 좌측에 예시된 새로운 raw_data_block에서, 콘텐츠는 스킵 가능한 블록으로 이루어질 수 있어서, 디코더는 관련 없는 부분을 스킵하고, 현재 디바이스에 대해 관련된 것으로 메타데이터에 의해 표시된 채널 요소만을 디코딩할 수 있다. 예에서, 스킵 가능한 블록은 raw_data_block 및 관련 정보를 포함한다.

[0080] 재전송을 위해 블록을 사용하는 것이 또한 가능할 것이다. 각각의 프레임은, 위의 스킵 가능한 블록 섹션에 설명된 바와 같이, 블록으로 분할될 수 있다. 블록은, 도 15 및 도 16에 예시된 바와 같이, 블록이 속한 프레임 번호, 동일한 ID를 갖는 상이한 프레임으로부터의 연속적인 블록을 블록 스트림에 연관시키기 위해 사용될 수 있는 블록 ID, 및 재전송을 위한 우선순위에 의해 식별될 수 있다. 예를 들어, 재전송에 대한 높은 우선순위(예를 들어, 우선순위 0으로 도 15 및 도 16에 표시됨)는 블록이 동일한 블록 ID 및 프레임 카운터를 갖지만 재전송에 대한 더 낮은 우선순위(예를 들어, 우선순위 1로 도 15 및 도 16에 표시됨)를 갖는 다른 블록에 비해 수신기에서 선호될 것임을 시그널링한다. 도 15 및 도 16의 예에서와 같이, 우선순위는 우선순위 인덱스가 증가함에 따라 감소될 수도 있는 반면(예를 들어, 우선순위 1은 우선순위 0보다 더 낮은 우선순위일 수도 있음), 다른 예에서, 우선순위는 우선순위 인덱스가 증가함에 따라 증가할 수도 있는 것에(예를 들어, 우선순위 0은 우선순위 1보다 더 낮은 우선순위임) 유의해야 한다. 또 다른 예가 당업자에게 명백할 것이다.

[0081] 신택스는 오디오 요소의 재전송을 지원할 수 있다. 재전송을 위해, 다양한 품질 레벨이 지원될 수 있다. 따라서 재전송된 블록은, 도 17 및 도 18에 예시된 바와 같이, 동일한 프레임 카운터 및 블록 ID로 수신된 블록이 중복이고 이에 따라 디코더에 대해 상호 배타적이기 때문에, 동일한 블록 ID를 갖는 어느 블록이 디코더에 대해 우선순위를 취해야 하는지를 표시하기 위해 '우선순위' 플래그를 운반할 수 있다.

[0082] 블록의 재전송은 감소된 데이터 레이트로 행해질 수 있다. 이러한 감소된 데이터 레이트는 오디오 신호의 신호 대 잡음 비를 감소시키고, 오디오 신호의 대역폭을 감소시키고, 오디오 신호의 채널 카운트를 감소시킴으로써 (예를 들어, 그 전체가 참조로 통합된 미국 특허 11,289,103에 설명된 바와 같이), 또는 이들의 임의의 조합에 의해 달성될 수 있다. 디코더가 최상의 가능한 품질을 제공하는 오디오 블록을 선택하기 위해, 최고 품질 신호를 제공하는 블록은 가장 높은 디코딩 우선순위를 가질 수 있고, 두 번째 높은 품질 신호를 갖는 블록은 두 번째 높은 우선순위를 가질 수 있는 식이다.

[0083] 블록은 또한 동일한 품질 레벨에서 재전송될 수 있다. 이러한 경우, 우선순위는 재전송된 블록의 레이턴시를 반영할 수 있다.

[0084] 코어 코딩 효율 개선을 위한 툴

[0085] 추가 예는 채널 요소를 통해 MDCT 양자화 스케일 팩터를 공유하는 것을 허용할 수 있다. 일부 경우에서, 연관된 사이드 비트 레이트를 감소시키기 위해 특정한 채널을 통해 MDCT 스케일 팩터를 공유하는 것이 가능할 수 있다.

또한, 스케일 팩터 공유는, 예를 들어, 7.1.4 입력의 경우처럼 상이한 채널 요소로 확장될 수 있다. 공유된 스케일 팩터의 사용은 세 개의 활성 구성을 허용하는 두 개의 시그널링 비트에 의해 표시될 수 있다. 하나의 가능한 구성은 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 및 우측 정상 채널에서 스케일 팩터를 공유하는 것일 것이다. 특정 신택스는, 스케일 팩터가 스킵 블록 내에서만 공유되는 것을 보장하기 위해, 스킵 가능한 블록 개념에 맞춰질 수 있다.

[0086] 둘 초과의 채널의 공동 코딩을 갖는 것이 또한 가능할 것이다. 도 4, 도 5, 도 6, 및 도 7에 예시된 예에서, 블록(Blk 1)은, 이 시나리오에서 고려된 스마트 스피커의 드라이버 각각에 대해 하나씩, 세 개의 SCE를 운반한다. 이러한 상황에서, (MPEG-D USAC로 지칭되는, 예를 들어 ISO/IEC 23003-3에 설명된 바와 같은 MDCT 기반 복잡한 예측 스테레오 툴인, 스테레오 예측을 포함하도록 확장될 수 있는, CPE에 의해 제공되는 바와 같은) 단지 두 개의 채널뿐만 아니라, 예를 들어, ETSI TS 103 190에 도입된 SAP(스테레오 오디오 프로세싱(Stereo Audio Processing)) 툴에 설명된 바와 같은 둘 초과의 채널의 공동 코딩을 가능하게 하는 것이 유익할 수 있다.

[0087] 유사하게, 가요성 렌더링의 경우, 디바이스들에 걸친 신호들 간의 높은 양의 상관이 있을 수 있다는 것에 유의한다. 이러한 신호에 대해, 다수의 디바이스를 커버하고, 위에서 개략된 툴을 사용하여 둘 초과의 채널을 통해 공동 채널 코딩이 적용되는 것을 허용하는 스킵 블록을 구성하는 것이 유익할 수 있다.

[0088] 특정한 변환 길이에 대해 사용될 수 있는 또 다른 공동 채널 코딩 툴은, 중간 및 고주파수에 대해 복합 채널 및 스케일 팩터 정보가 전송되는 채널 커플링이다. 이는 양호한 품질 범위의 재생을 위해 비트레이트 감소를 제공할 수 있다. 예를 들어, 낮은 레이턴시 코딩을 위해 256 샘플 주위의 프레임 길이에 대응하는, 256의 변환 길이를 갖는 프레임에 대한 채널 커플링 툴을 사용하는 것이 유익할 수 있다.

[0089] 본 개시는 대역제한된 신호의 효율적인 코딩을 또한 허용할 것이다. 스마트 스피커의 상이한 드라이버를 피드하는 별개의 신호가 송신되는 시나리오에서, 이러한 신호 중 일부는, 예를 들어, 우퍼(woofer), 중간 범위, 및 트위터(tweeter)를 갖는 3-방 드라이버 구성의 경우와 같이, 대역제한될 수 있다. 그러므로, 이러한 대역제한된 신호의 효율적인 인코딩이 바람직하며, 이는 개선된 코딩 효율 및/또는 감소된 계산 복잡도(예를 들어, 우퍼 피드에 대한 대역제한된 IMDCT의 사용을 가능하게 함)로 이러한 시나리오를 처리하기 위해 특별히 튜닝된 심리음향 모델 및 비트 할당 전략뿐만 아니라 신택스의 잠재적인 수정으로 바뀔 수 있다.

[0090] 일부 예에서, MDCT 도메인에서의 고주파 재구성 및 잡음 추가의 결합이 가능할 것이다. 현대 오디오 코덱은 통상적으로 파라메트릭 코딩 기술을 지원하도록 설계되고, 유사하게 여기서, 낮은 레이턴시를 유지하고 재구성된 고대역에서 톤 대 잡음비를 관리하기 위해 MDCT 도메인에서 잡음 추가 체계를 수행하기 위해, 예를 들어, MDCT 도메인에서 고주파수 재구성 방법을 포함하는 것을 상상할 수 있다. 이러한 파라메트릭 코딩 기술은 특히 더 낮은 동작 포인트에 대해, 그리고 특히 재전송이 FFC(순방향 에러 정정(Forward Error Correction)) 체계의 부분으로서 발생하는 시나리오에서 유용할 수 있고, 여기서 통상적으로 메인 신호가 전송되고, 그 후 동일한 신호가 지연된 방식으로 더 낮은 비트레이트로 재전송된다.

[0091] 이러한 예는 인코더의 피크 복잡도 감소를 추가로 허용할 것이다. 이는 일정한 비트레이트 전달 채널에 대한 버퍼 모델 제한이 없는 경우에 적용될 수 있을 것이다. 버퍼 모드를 갖는 일정한 비트레이트에서, 양자화 및 카운팅 후에, 인코딩될 프레임에 대해 사용될 비트의 결과적인 수가 허용된 비트의 한도 위인 것이 발생할 수 있다. 적어도 하나의 새로운, 보다 조악한 양자화 및 비트 카운팅 단계가 비트 요건을 충족시키기 위해 수행될 필요가 있다. 이 한도가 보다 완화되면, 인코더는 처음 양자화 결과를 유지할 수 있고 약간 더 높은 순간 비트레이트로 행해질 것이다. 버퍼 모델 인코더를 갖는 일정한 비트레이트의 거동과 유사한 인코더 거동은, 버퍼 요건을 따르는 인코더에 따라 업데이트되는 가상 버퍼 충만도와 함께, 동일한 비트 저장소 제어 메커니즘을 사용함으로써 여전히 달성될 수 있다. 이는 추가적인 양자화 및 비트 카운팅 단계를 줄일 것이며, 결과적인 오디오 품질은 약간 증가된 전체 비트레이트의 단점으로 일정한 비트레이트 경우와 비교하여 동일하거나 더 양호할 것이다.

[0092] 리턴 채널 개념

[0093] 재생 및 청취 사용-사례의 경우, 스마트 스피커(30)는 (예를 들어, 모노, 스테레오, A-포맷, B-포맷, 또는 임의의 다른 등방성 또는 이방성 채널 포맷의) 모노 또는 공간 음장(spatial soundfield)을 캡처하도록 구성되는 하나 이상의 마이크로폰(40)(예를 들어, 마이크로폰 또는 마이크로폰 어레이)을 가질 수 있고, 코덱이 낮은 레이턴시 방식으로 이러한 포맷의 효율적인 코딩을 행하는 것이 가능할 필요가 있어서, 예를 들어, 도 12에 예시된, 리턴 채널에 대해 사용되는 바와 같이 동일한 코덱이 스마트 스피커 디바이스(30)로의 브로드캐스트/전송을 위해 사용된다.

[0094] 웨이크 워드 검출이 스마트 스피커 디바이스 상에서 수행되는 동안, 음성 인식은 통상적으로, 웨이크 워드 검출에 의해 트리거된 기록된 음성의 적절한 세그먼트에 의해, 클라우드에서 행해진다는 것에 또한 주의해야 한다.

[0095] 이러한 맥락에서, 예를 들어, 코덱 내의 중간 포맷/표현 상에 음성 분석 시스템을 생성함으로써 시스템 복잡도를 전반적으로 줄이는 것에 관심을 가질 수 있다. 사람-대-사람 대화가 진행되지 않는 가장 간단한 사용-사례의 경우, 적합한 표현은 다른 사람에 의해 청취되지 않기 때문에 음성 인식 태스크에 명시적으로 정의될 수 있다. 이는 대역 에너지, Mel-frequency Cepstral Coefficients(MFCC) 등, 또는 MDCT 코딩된 스펙트럼의 낮은 비트레이트 버전일 수 있다.

[0096] 진행중인 사람-대-사람의 대화가 있고, 웨이크 워드와 요구되는 음성 인식이 중간에 삽입된 사용-사례에서, 사람은 그것을 위해 스트림의 관련 부분을 추출하는 것이 가능하기를 원할 것이다. 이러한 추출은 메인 음성 신호에 병렬로 송신되는 단순히 추가적인 데이터의 계층화된 코딩 구조를 수반할 수 있다. 또한 기존의 디코딩된 MDCT 스펙트럼으로부터 관련된 표현으로 정의된 트랜스코드를 상상하고, 그렇게 함으로써 구조를 단순화한다. 본질적으로 디코더는 음성 인식 표현을 또한 (병렬로) 디코딩해야 한다고 시그널링될 때까지 사람 청취 가능 오디오를 디코딩 및 출력할 것이다. 이 표현은 "박리된" 계층화된 스트림, 단순히 동일한 완전한 데이터의 대안적인 디코드, 또는 단순히 스트림 내의 추가적인 표현의 디코드 및 출력일 수 있다. 이 사용-사례에서, 시그널링은 수신측의 디코더가 음성 인식 관련 표현을 출력해야 한다는 것을 표시하는 활성화 피스이다.

[0097] 열거된 예

[0098] 이하에서, 청구항이 아닌 열거된 예시적인 예의 7개의 세트들(EEE-A, EEE-B, EEE-C, EEE-D, EEE-E, EEE-F, 및 EEE-G)은 본원에 개시된 예의 양태를 설명한다.

[0099] EEE-A1. 오디오 신호를 디코딩하기 위한 방법으로서,

[0100] 적어도 하나의 프레임을 포함하는 비트스트림을 수신하는 단계 - 적어도 하나의 프레임의 각각의 프레임은 복수의 블록을 포함함 -;

[0101] 출력 디바이스의 디바이스 정보에 기초하여, 시그널링 데이터로부터, 디코딩 시에 스킵될 복수의 블록 중 하나 이상의 블록의 부분을 식별하기 위한 정보를 결정하는 단계; 및

[0102] 하나 이상의 블록의 식별된 부분을 스킵하면서 비트스트림을 디코딩하는 단계를 포함하는, 방법.

[0103] EEE-A2. EEE-A1의 방법에 있어서, 디코딩 시에 스킵될 복수의 블록 중 하나 이상의 블록의 부분을 식별하기 위한 정보는 복수의 출력 디바이스의 각각의 출력 디바이스를 하나 이상의 비트스트림 요소에 연관시키는 매트릭스를 포함하는, 방법.

[0104] EEE-A3. EEE-A2의 방법에 있어서, 하나 이상의 비트스트림 요소는 대응하는 연관된 출력 디바이스에 대한 비트스트림의 디코딩을 위해 요구되는, 방법.

[0105] EEE-A4. EEE-A1 내지 EEE-A3 중 어느 하나의 방법에 있어서, 출력 디바이스는 무선 디바이스, 모바일 디바이스, 태블릿, 단일 채널 스피커, 및/또는 멀티 채널 스피커 중 적어도 하나를 포함할 수 있는, 방법.

[0106] EEE-A5. EEE-A1 내지 EEE-A4 중 어느 하나의 방법에 있어서, 식별된 부분은 적어도 하나의 블록을 포함하는, 방법.

[0107] EEE-A6. EEE-A1 내지 EEE-A5 중 어느 하나의 방법에 있어서, 출력 디바이스는 제1 출력 디바이스이고, 비트스트림의 하나 이상의 신호들 간의 공동 코딩 기술을 제2 출력 디바이스 및 제3 출력 디바이스에 적용하는 단계를 더 포함하는, 방법.

[0108] EEE-A7. EEE-A1 내지 EEE-A6 중 어느 하나의 방법에 있어서, 각각의 출력 디바이스 및/또는 디코더의 아이덴티티가 시스템 초기화 페이즈 동안 정의되는, 방법.

[0109] EEE-A8. EEE-A1 내지 EEE-A7 중 어느 하나의 방법에 있어서, 시그널링 데이터는 비트스트림의 메타데이터로부터 결정되는, 방법.

[0110] EEE-A9. EEE-A1 내지 EEE-A8 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0111] EEE-A10. 실행될 때, 하나 이상의 디바이스가 EEE-A1 내지 EEE-A8 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0112] EEE-B1. 복수의 오디오 신호를 포함하는 오디오 프로그램으로부터 인코딩된 비트스트림을 생성하기 위한 방법으로서,

[0113] 복수의 오디오 신호 각각에 대해, 개개의 오디오 신호가 연관된 재생 디바이스를 표시하는 정보를 수신하는 단계;

[0114] 각각의 재생 디바이스에 대해, 개개의 재생 디바이스와 연관된 지연, 이득, 및 동등화 곡선 중 적어도 하나를 표시하는 정보를 수신하는 단계;

[0115] 복수의 오디오 신호로부터, 둘 이상의 관련된 오디오 신호의 그룹을 결정하는 단계;

[0116] 공동-코딩된 오디오 신호를 획득하기 위해 하나 이상의 공동-코딩 툴을 그룹의 둘 이상의 관련된 오디오 신호에 적용하는 단계;

[0117] 공동으로 코딩된 오디오 신호, 공동으로 코딩된 오디오 신호가 연관되는 재생 디바이스의 표시, 및 공동으로 코딩된 오디오 신호가 연관되는 개개의 재생 디바이스와 연관된 지연 및 이득의 표시를, 인코딩된 비트스트림의 독립적인 블록에 결합하는 단계를 포함하는, 방법.

[0118] EEE-B2. EEE-B1의 방법에 있어서, 개개의 재생 디바이스와 연관된 지연, 이득, 및/또는 동등화 곡선은 청취자의 위치에 대한 개개의 재생 디바이스의 위치에 의존하는, 방법.

[0119] EEE-B3. EEE-B1 또는 EEE-B2의 방법에 있어서, 개개의 재생 디바이스와 연관된 지연, 이득, 및/또는 동등화 곡선은 다른 재생 디바이스의 위치에 대한 개개의 재생 디바이스의 위치에 의존하는, 방법.

[0120] EEE-B4. EEE-B1 내지 EEE-B3 중 어느 하나의 방법에 있어서, 지연, 이득, 및/또는 동등화 곡선은 동적으로 가변적인, 방법.

[0121] EEE-B5. EEE-B4의 방법에 있어서, 지연, 이득, 및/또는 동등화 곡선은 청취자의 위치의 변경에 응답하여 조정되는, 방법.

[0122] EEE-B6. EEE-B4 또는 EEE-B5의 방법에 있어서, 지연, 이득, 및/또는 동등화 곡선은 재생 디바이스의 위치의 변경에 응답하여 조정되는, 방법.

[0123] EEE-B7. EEE-B4 내지 EEE-B6 중 어느 하나의 방법에 있어서, 지연, 이득, 및/또는 동등화 곡선은 다른 재생 디바이스 중 하나 이상의 위치의 변경에 응답하여 조정되는, 방법.

[0124] EEE-B8. EEE-B1 내지 EEE-B7 중 어느 하나의 방법에 있어서, 복수의 오디오 신호로부터, 둘 이상의 관련된 오디오 신호의 그룹의 일부가 아닌 오디오 신호를 결정하는 단계를 더 포함하는, 방법.

[0125] EEE-B9. EEE-B8의 방법에 있어서, 둘 이상의 관련된 오디오 신호의 그룹의 일부가 아닌 오디오 신호에 대해, 오디오 신호가 연관되는 재생 디바이스와 연관된 지연, 이득, 및/또는 동등화 곡선을 적용하는 단계를 더 포함하는, 방법.

[0126] EEE-B10. EEE-B9의 방법에 있어서, 둘 이상의 관련된 오디오 신호의 그룹의 일부가 아닌 오디오 신호를 독립적으로 코딩하는 단계, 및 독립적으로 코딩된 오디오 신호 및 독립적으로 코딩된 오디오 신호가 연관되는 재생 디바이스의 표시를 인코딩된 비트스트림의 별개의 독립적으로 디코딩 가능한 서브세트에 결합하는 단계를 더 포함하는, 방법.

[0127] EEE-B11. EEE-B8의 방법에 있어서, 둘 이상의 관련된 오디오 신호의 그룹의 일부가 아닌 오디오 신호를 독립적으로 코딩하는 단계, 및 독립적으로 코딩된 오디오 신호, 독립적으로 코딩된 오디오 신호가 연관되는 재생 디바이스의 표시, 및 독립적으로 코딩된 오디오 신호가 연관되는 재생 디바이스와 연관되는 지연, 이득, 및/또는 동등화 곡선의 표시를 인코딩된 비트스트림의 별개의 독립적으로 디코딩 가능한 서브세트에 결합하는 단계를 더 포함하는, 방법.

[0128] EEE-B12. 인코딩된 비트스트림의 프레임으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은:

[0129] 인코딩된 비트스트림으로부터, 재생 디바이스와 연관된 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 독립적인 블록을 식별하는 단계;

[0130] 인코딩된 비트스트림으로부터, 인코딩된 데이터의 식별된 독립적인 블록을 추출하는 단계;

[0131] 인코딩된 데이터의 추출된 독립적인 블록이 둘 이상의 공동으로 코딩된 오디오 신호를 포함한다고 결정하는 단계;

[0132] 재생 디바이스와 연관된 하나 이상의 오디오 신호를 획득하기 위해 둘 이상의 공동-코딩된 오디오 신호에 하나 이상의 공동-디코딩 툴을 적용하는 단계;

[0133] 인코딩된 데이터의 추출된 독립적인 블록으로부터, 재생 디바이스와 연관된 지연, 이득, 및 동등화 곡선 중 적어도 하나를 결정하는 단계;

[0134] 재생 디바이스와 연관된 지연, 이득, 및/또는 동등화 곡선을 재생 디바이스와 연관된 하나 이상의 오디오 신호에 적용하는 단계를 포함하는, 방법.

[0135] EEE-B13. EEE-B12의 방법에 있어서, 재생 디바이스와 연관된 결정된 지연, 이득, 및/또는 동등화 곡선은 청취자의 위치에 대한 재생 디바이스의 위치에 의존하는, 방법.

[0136] EEE-B14. EEE-B12 또는 EEE-B13의 방법에 있어서, 재생 디바이스와 연관된 결정된 지연, 이득, 및/또는 동등화 곡선은 다른 재생 디바이스에 대한 재생 디바이스의 위치에 의존하는, 방법.

[0137] EEE-B15. EEE-B12 내지 EEE-B14 중 어느 하나의 방법에 있어서, 재생 디바이스의 결정된 지연, 이득, 및/또는 동등화 곡선은 동적으로 가변적인, 방법.

[0138] EEE-B16. EEE-B15의 방법에 있어서, 재생 디바이스와 연관된 결정된 지연, 이득, 및/또는 동등화 곡선이 재생 디바이스와 연관된 이전에 결정된 지연, 이득, 및/또는 동등화 곡선과 상이할 때, 방법은 재생 디바이스와 연관된 이전에 결정된 지연, 이득, 및/또는 동등화 곡선과 재생 디바이스와 연관된 결정된 지연, 이득, 및/또는 동등화 곡선 사이에서 보간하는 단계를 더 포함하는, 방법.

[0139] EEE-B17. EEE-B16의 방법에 있어서, 결정된 지연, 이득, 및/또는 동등화 곡선은 청취자의 위치의 변경으로 인해 이전에 결정된 지연, 이득, 및/또는 동등화 곡선과 상이한, 방법.

[0140] EEE-B18. EEE-B16 또는 EEE-B17의 방법에 있어서, 결정된 지연, 이득, 및/또는 동등화 곡선은 재생 디바이스의 위치의 변경으로 인해 이전에 결정된 지연, 이득, 및/또는 동등화 곡선과 상이한, 방법.

[0141] EEE-B19. EEE-B16 내지 EEE-B18 중 어느 하나의 방법에 있어서, 결정된 지연, 이득, 및/또는 동등화 곡선은 다른 재생 디바이스 중 하나 이상의 위치의 변경으로 인해 이전에 결정된 지연, 이득, 또는 동등화 곡선과 상이한, 방법.

[0142] EEE-B20. EEE-B12 내지 EEE-B19 중 어느 하나의 방법에 있어서, 인코딩된 비트스트림의 프레임은 인코딩된 데이터의 둘 이상의 독립적인 블록을 포함하고, 방법은:

[0143] 독립적인 블록 중 하나 이상이 재생 디바이스와 연관되지 않은 오디오 신호를 포함한다고 결정하는 단계; 및

[0144] 재생 디바이스와 연관되지 않은 오디오 신호를 포함하는 하나 이상의 독립적인 블록을 무시하는 단계를 더 포함하는, 방법.

[0145] EEE-B21. EEE-B12 내지 EEE-B20 중 어느 하나의 방법에 있어서, 하나 이상의 공동-디코딩 툴을 적용하는 단계는 재생 디바이스와 연관된 공동-코딩된 오디오 신호의 서브세트를 식별하는 단계, 및 재생 디바이스와 연관된 하나 이상의 오디오 신호를 획득하기 위해 공동-코딩된 오디오 신호의 그 서브세트만을 재구성하는 단계를 포함하는, 방법.

[0146] EEE-B22. EEE-B12 내지 EEE-B20 중 어느 하나의 방법에 있어서, 하나 이상의 공동-디코딩 툴을 적용하는 단계는 공동-코딩된 오디오 신호 각각을 재구성하는 단계, 재생 디바이스가 연관된 재구성된 공동-코딩된 오디오 신호의 서브세트를 식별하는 단계, 및 재생 디바이스와 연관된 재구성된 공동-코딩된 오디오 신호의 서브세트로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 획득하는 단계를 포함하는, 방법.

[0147] EEE-B23. EEE-B1 내지 EEE-B22 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0148] EEE-B24. 실행될 때, 하나 이상의 디바이스가 EEE-B1 내지 EEE-B22 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0149] EEE-C1. 복수의 오디오 신호를 포함하는 오디오 프로그램의 인코딩된 비트스트림의 프레임을 생성하기 위한 방법으로서, 프레임은 인코딩된 데이터의 둘 이상의 독립적인 블록을 포함하고, 방법은:

[0150] 복수의 오디오 신호 중 하나 이상에 대해, 하나 이상의 오디오 신호가 연관되는 재생 디바이스를 표시하는 정보를 수신하는 단계;

[0151] 표시된 재생 디바이스에 대해, 하나 이상의 추가적인 연관된 재생 디바이스를 표시하는 정보를 수신하는 단계;

[0152] 표시된 하나 이상의 추가적인 연관된 재생 디바이스와 연관된 하나 이상의 오디오 신호를 수신하는 단계;

[0153] 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계;

[0154] 표시된 하나 이상의 추가적인 연관된 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계;

[0155] 재생 디바이스와 연관된 하나 이상의 인코딩된 오디오 신호 및 하나 이상의 추가적인 연관된 재생 디바이스를 표시하는 시그널링 정보를 제1 독립적인 블록에 결합하는 단계;

[0156] 하나 이상의 추가적인 연관된 재생 디바이스와 연관된 하나 이상의 인코딩된 오디오 신호를 하나 이상의 추가적인 독립적인 블록에 결합하는 단계; 및

[0157] 제1 독립적인 블록 및 하나 이상의 추가적인 독립적인 블록을 인코딩된 비트스트림의 프레임에 결합하는 단계를 포함하는, 방법.

[0158] EEE-C2. EEE-C1의 방법에 있어서, 복수의 오디오 신호는 재생 디바이스 또는 하나 이상의 추가적인 연관된 재생 디바이스와 연관되지 않은 오디오 신호의 하나 이상의 그룹을 포함하고,

[0159] 재생 디바이스 또는 하나 이상의 추가적인 연관된 재생 디바이스와 연관되지 않은 오디오 신호의 하나 이상의 그룹 각각을 개개의 독립적인 블록으로 인코딩하는 단계; 및

[0160] 하나 이상의 그룹 각각에 대한 개개의 독립적인 블록을 인코딩된 비트스트림의 프레임에 결합하는 단계를 더 포함하는, 방법.

[0161] EEE-C3. EEE-C1 또는 EEE-C2의 방법에 있어서, 표시된 하나 이상의 추가적인 연관된 재생 디바이스와 연관된 하나 이상의 오디오 신호는 재생 디바이스에 대한 에코-관리를 수행하기 위한 에코-참조로서 사용되도록 특별히 의도되는, 방법.

[0162] EEE-C4. EEE-C3의 방법에 있어서, 에코-참조로서 사용되도록 의도된 하나 이상의 오디오 신호는 재생 디바이스와 연관된 하나 이상의 오디오 신호보다 적은 데이터를 사용하여 전송되는, 방법.

[0163] EEE-C5. EEE-C3 또는 EEE-C4의 방법에 있어서, 하나 이상의 오디오 신호는 파라메트릭 코딩 툴을 사용하여 인코딩된 에코-참조로서 사용되도록 의도되는, 방법.

[0164] EEE-C6. EEE-C1 또는 EEE-C2의 방법에 있어서, 하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 오디오 신호는 하나 이상의 다른 재생 디바이스로부터의 재생에 적합한, 방법.

[0165] EEE-C7. 인코딩된 비트스트림의 프레임으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하기 위한 방법으로서, 프레임은 인코딩된 데이터의 둘 이상의 독립적인 블록을 포함하고, 재생 디바이스는 하나 이상의 마이크로폰을 포함하고, 방법은:

[0166] 인코딩된 비트스트림으로부터, 재생 디바이스와 연관된 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 독립적인 블록을 식별하는 단계;

[0167] 인코딩된 비트스트림으로부터, 인코딩된 데이터의 식별된 독립적인 블록을 추출하는 단계;

[0168] 인코딩된 데이터의 식별된 독립적인 블록으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 추출하는 단계;

[0169] 인코딩된 비트스트림으로부터, 하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 하나 이상의 다른 독립적인 블록을 식별하는 단계;

[0170] 인코딩된 데이터의 하나 이상의 다른 독립적인 블록으로부터 하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 오디오 신호를 추출하는 단계;

[0171] 재생 디바이스의 하나 이상의 마이크로폰을 사용하여 하나 이상의 오디오 신호를 캡처하는 단계; 및

[0172] 하나 이상의 캡처된 오디오 신호에 응답하여 재생 디바이스에 대한 에코-관리를 수행하기 위한 에코-참조로서

하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 추출된 오디오 신호를 사용하는 단계를 포함하는, 방법.

[0173] EEE-C8. EEE-C7의 방법에 있어서,

[0174] 인코딩된 비트스트림이 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 포함한다고 결정하는 단계; 및

[0175] 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 무시하는 단계를 더 포함하는, 방법.

[0176] EEE-C9. EEE-C8의 방법에 있어서, 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 무시하는 단계는 인코딩된 데이터의 추가적인 하나 이상의 독립적인 블록을 추출하지 않고 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 스킵하는 단계를 포함하는, 방법.

[0177] EEE-C10. EEE-C7 내지 EEE-C9 중 어느 하나의 방법에 있어서, 하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 오디오 신호는 재생 디바이스에 대한 에코-관리를 수행하기 위한 에코-참조로서 사용되도록 특별히 의도되는, 방법.

[0178] EEE-C11. EEE-C10의 방법에 있어서, 에코-참조로서 사용되도록 특별히 의도된 하나 이상의 오디오 신호는 재생 디바이스와 연관된 하나 이상의 오디오 신호보다 적은 데이터를 사용하여 전송되는, 방법.

[0179] EEE-C12. EEE-C10 또는 EEE-C11의 방법에 있어서, 에코-참조로서 사용되도록 특별히 의도된 하나 이상의 오디오 신호는 하나 이상의 오디오 신호의 파라메트릭 표현으로부터 재구성되는, 방법.

[0180] EEE-C13. EEE-C7 내지 EEE-C9 중 어느 하나의 방법에 있어서, 하나 이상의 다른 재생 디바이스와 연관된 하나 이상의 오디오 신호는 하나 이상의 다른 재생 디바이스로부터의 재생에 적합한, 방법.

[0181] EEE-C14. EEE-C7의 방법에 있어서, 인코딩된 신호는 재생 디바이스에 대한 에코-참조로서 사용하기 위해 하나 이상의 다른 재생 디바이스를 표시하는 시그널링 정보를 포함하는, 방법.

[0182] EEE-C15. EEE-C14의 방법에 있어서, 현재 프레임에 대한 시그널링 정보에 의해 표시된 하나 이상의 다른 재생 디바이스는 이전 프레임에 대한 에코-참조로서 사용된 하나 이상의 다른 재생 디바이스와 상이한, 방법.

[0183] EEE-C16. EEE-C1 내지 EEE-C15 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0184] EEE-C17. 실행될 때, 하나 이상의 디바이스가 EEE-C1 내지 EEE-C15 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0185] EEE-D1. 오디오 신호를 전송하기 위한 방법으로서,

[0186] 비트스트림의 부분을 포함하는 데이터의 패킷을 생성하는 단계 - 비트스트림은 복수의 프레임을 포함하고, 복수의 프레임의 각각의 프레임은 복수의 블록을 포함하고, 데이터의 패킷을 생성하는 단계는:

[0187] 데이터의 패킷과 복수의 블록의 하나 이상의 블록을 어셈블링하는 단계를 포함하고, 상이한 프레임으로부터의 블록은 단일 패킷에 결합되고 및/또는 비순차적으로 전송됨 -; 및

[0188] 패킷 기반 네트워크를 통해 데이터 패킷을 전송하는 단계를 포함하는, 방법.

[0189] EEE-D2. EEE-D1의 방법에 있어서, 복수의 블록의 각각의 블록은 식별 정보를 포함하는, 방법.

[0190] EEE-D3. EEE-D2의 방법에 있어서, 식별 정보는 블록 ID, 블록과 연관된 대응하는 프레임 번호, 및/또는 재전송을 위한 우선순위 중 적어도 하나를 포함하는, 방법.

[0191] EEE-D4. EEE-D1 내지 EEE-D3 중 어느 하나의 방법에 있어서, 복수의 프레임의 각각의 프레임은 시작 시간, 종료 시간, 및 지속기간을 갖는 오디오 신호의 연속적인 세그먼트를 표현하는 모든 오디오 데이터를 운반하는, 방법.

[0192] EEE-D5. 오디오 신호를 디코딩하기 위한 방법으로서,

[0193] 비트스트림의 부분을 포함하는 데이터의 패킷을 수신하는 단계 - 비트스트림은 복수의 프레임을 포함하고, 복수의 프레임의 각각의 프레임은 복수의 블록을 포함함 -;

[0194] 복수의 블록 중 디바이스로 어드레싱된 블록의 세트를 결정하는 단계; 및

[0195] 디바이스로 어드레싱된 블록의 세트를 디코딩하고 복수의 블록 중 디바이스로 어드레싱되지 않은 블록의 디코딩을 스킵하는 단계를 포함하는, 방법.

[0196] EEE-D6. 오디오 스트림을 전송하기 위한 방법으로서,

[0197] 오디오 스트림을 전송하는 단계를 포함하고, 오디오 스트림은 복수의 프레임을 포함하고, 복수의 프레임의 각각의 프레임은 복수의 블록을 포함하고, 전송하는 단계는 오디오 스트림에 대한 구성 정보를 대역 외로 전송하는 단계를 포함하는, 방법.

[0198] EEE-D7. EEE-D6의 방법에 있어서, 오디오 스트림에 대한 구성 정보를 대역 외로 전송하는 단계는:

[0199] 제1 네트워크 및/또는 제1 네트워크 프로토콜을 통해 오디오 스트림을 전송하는 단계; 및

[0200] 제2 네트워크 및/또는 제2 네트워크 프로토콜을 통해 구성 정보를 전송하는 단계를 포함하는, 방법.

[0201] EEE-D8. EEE-D7의 방법에 있어서, 제1 네트워크 프로토콜은 사용자 데이터그램 프로토콜(UDP)이고, 제2 네트워크 프로토콜은 전송 제어 프로토콜(TCP)인, 방법.

[0202] EEE-D9. 오디오 신호를 디코딩하기 위한 방법으로서,

[0203] 비트스트림을 수신하는 단계 - 비트스트림은:

[0204] 정적 구성 양태의 시그널링에 대응하는 정보;

[0205] 정적 메타데이터를 포함함 -; 및

[0206] 정보 및/또는 정적 메타데이터에 기초하여 하나 이상의 채널 요소를 하나 이상의 디바이스에 매핑하는 단계를 포함하는, 방법.

[0207] EEE-D10. EEE-D9의 방법에 있어서, 비트스트림은 비트스트림을 디코딩하도록 구성된 복수의 디코더에 의해 수신되고, 복수의 디코더의 각각의 디코더는 비트스트림의 부분을 디코딩하도록 구성되는, 방법.

[0208] EEE-D11. EEE-D9 또는 EEE-D10의 방법에 있어서, 비트스트림은 동적 메타데이터를 더 포함하는, 방법.

[0209] EEE-D12. EEE-D9 내지 EEE-D11 중 어느 하나의 방법에 있어서, 비트스트림은 복수의 블록을 포함하고, 복수의 블록의 각각의 블록은:

[0210] 디코딩 동안 블록의 부분이 스킵되는 것을 가능하게 하는 정보 - 부분은 디바이스에 대해 필요하지 않음 -; 및

[0211] 동적 메타데이터를 포함하는, 방법.

[0212] EEE-D13. 오디오 신호의 블록을 재전송하기 위한 방법으로서,

[0213] 비트스트림의 하나 이상의 블록을 전송하는 단계를 포함하며, 비트스트림은 복수의 블록을 포함하고, 비트스트림의 하나 이상의 블록 각각은 이전에 전송되었었고; 및

[0214] 하나 이상의 블록 각각은 디코딩 우선순위 표시자를 포함하는, 방법.

[0215] EEE-D14. EEE-D13의 방법에 있어서, 디코딩 우선순위 표시자는 비트스트림의 하나 이상의 블록을 디코딩하기 위한 우선순위의 순서를 디코더에 표시하는, 방법.

[0216] EEE-D15. EEE-D13 또는 EEE-D14의 방법에 있어서, 하나 이상의 블록의 각각의 블록은 동일한 블록 ID를 포함하는, 방법.

[0217] EEE-D16. EEE-D13 내지 EEE-D15 중 어느 하나의 방법에 있어서, 비트스트림의 하나 이상의 블록의 전송은 이전의 전송과 비교하여 데이터 레이트를 감소시킴으로써 전송되는, 방법.

[0218] EEE-D17. EEE-D16의 방법에 있어서, 데이터 레이트를 감소시키는 것은 오디오 신호의 신호 대 잡음 비를 감소시키는 것, 오디오 신호의 대역폭을 감소시키는 것, 및/또는 오디오 신호의 채널 카운트를 감소시키는 것 중 적어도 하나를 포함하는, 방법.

[0219] EEE-E1. 복수의 오디오 신호를 포함하는 오디오 프로그램의 인코딩된 비트스트림의 프레임을 생성하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은:

[0220] 복수의 오디오 신호 각각에 대해, 개개의 오디오 신호가 연관된 재생 디바이스를 표시하는 정보를 수신하는 단계;

[0221] 하나 이상의 인코딩된 오디오 신호를 획득하기 위해 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를

인코딩하는 단계;

[0222] 개개의 재생 디바이스와 연관된 하나 이상의 인코딩된 오디오 신호를 프레임의 제1 독립적인 블록에 결합하는 단계;

[0223] 복수의 오디오 신호 중 하나 이상의 다른 오디오 신호를 하나 이상의 추가적인 독립적인 블록으로 인코딩하는 단계; 및

[0224] 제1 독립적인 블록 및 하나 이상의 추가적인 독립적인 블록을 인코딩된 비트스트림의 프레임에 결합하는 단계를 포함하는, 방법.

[0225] EEE-E2. EEE-E1의 방법에 있어서, 둘 이상의 오디오 신호는 재생 디바이스와 연관되고, 둘 이상의 오디오 신호 각각은 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된 신호이고, 대역제한된 신호 각각에 대해 상이한 인코딩 기술이 사용되는, 방법.

[0226] EEE-E3. EEE-E2의 방법에 있어서, 대역제한된 신호 각각에 대해 상이한 심리음향 모델 및/또는 상이한 비트 할당 기술이 사용되는, 방법.

[0227] EEE-E4. EEE-E1 내지 EEE-E3 중 어느 하나의 방법에 있어서, 인코딩된 신호의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델에 의해 제약되는, 방법.

[0228] EEE-E5. EEE-E1의 방법에 있어서, 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호를 인코딩하는 단계는 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 프레임의 제1 독립적인 블록으로 공동으로 인코딩하는 단계를 포함하는, 방법.

[0229] EEE-E6. EEE-E5의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 이상의 오디오 신호에 걸쳐 하나 이상의 스케일 팩터를 공유하는 단계를 포함하는, 방법.

[0230] EEE-E7. EEE-E6의 방법에 있어서, 둘 이상의 오디오 신호는 공간적으로 관련되는, 방법.

[0231] EEE-E8. EEE-E7의 방법에 있어서, 둘 이상의 공간적으로 관련된 오디오 신호는 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 또는 우측 정상 채널을 포함하는, 방법.

[0232] EEE-E9. EEE-E5의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 커플링 툴을 적용하는 단계를 포함하며, 커플링 툴을 적용하는 단계는:

[0233] 둘 이상의 오디오 신호를 명시된 주파수 위의 복합 신호로 결합하는 단계; 및

[0234] 둘 이상의 오디오 신호 각각에 대해, 복합 신호의 에너지 및 각각의 개개의 신호의 에너지에 관한 스케일 팩터를 결정하는 단계를 포함하는, 방법.

[0235] EEE-E10. EEE-E5의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 인코딩하는 단계는 둘 초과의 신호에 공동-코딩 툴을 적용하는 단계를 포함하는, 방법.

[0236] EEE-E11. 인코딩된 비트스트림의 프레임으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하기 위한 방법으로서, 프레임은 인코딩된 데이터의 하나 이상의 독립적인 블록을 포함하고, 방법은:

[0237] 인코딩된 비트스트림으로부터, 재생 디바이스와 연관된 하나 이상의 오디오 신호에 대응하는 인코딩된 데이터의 독립적인 블록을 식별하는 단계;

[0238] 인코딩된 비트스트림으로부터, 인코딩된 데이터의 식별된 독립적인 블록을 추출하는 단계;

[0239] 하나 이상의 디코딩된 오디오 신호를 획득하기 위해 인코딩된 데이터의 독립적인 블록으로부터 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하는 단계;

[0240] 인코딩된 비트스트림으로부터, 하나 이상의 추가적인 오디오 신호에 대응하는 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 식별하는 단계; 및

[0241] 인코딩된 데이터의 하나 이상의 추가적인 독립적인 블록을 디코딩하거나 스킵하는 단계를 포함하는, 방법.

[0242] EEE-E12. EEE-E11의 방법에 있어서, 둘 이상의 오디오 신호는 재생 디바이스와 연관되고, 둘 이상의 오디오 신호 각각은 재생 디바이스의 개개의 드라이버에 의한 재생을 위해 의도된 대역제한된 신호이고, 상이한 디코딩

기술이 둘 이상의 오디오 신호를 디코딩하는 데 사용되는, 방법.

[0243] EEE-E13. EEE-E12의 방법에 있어서, 상이한 심리음향 모델 및/또는 상이한 비트 할당 기법이 대역제한된 신호 각각을 인코딩하기 위해 사용된, 방법.

[0244] EEE-E14. EEE-E11 또는 EEE-E13의 방법에 있어서, 인코딩된 비트스트림의 순간 프레임 레이트는 가변적이고, 버퍼 충만도 모델에 의해 제약되는, 방법.

[0245] EEE-E15. EEE-E11의 방법에 있어서, 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하는 단계는 개개의 재생 디바이스와 연관된 하나 이상의 오디오 신호 및 하나 이상의 추가적인 재생 디바이스와 연관된 하나 이상의 추가적인 오디오 신호를 인코딩된 데이터의 독립적인 블록으로부터 공동으로 디코딩하는 단계를 포함하는, 방법.

[0246] EEE-E16. EEE-E15의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 둘 이상의 오디오 신호에 걸쳐 공유되는 스케일 팩터를 추출하는 단계를 포함하는, 방법.

[0247] EEE-E17. EEE-E16의 방법에 있어서, 둘 이상의 오디오 신호는 공간적으로 관련되는, 방법.

[0248] EEE-E18. EEE-E17의 방법에 있어서, 둘 이상의 공간적으로 관련된 오디오 신호는 좌측 수평 채널, 좌측 정상 채널, 우측 수평 채널, 또는 우측 정상 채널을 포함하는, 방법.

[0249] EEE-E19. EEE-E15의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 디커플링 툴을 적용하는 단계를 포함하는, 방법.

[0250] EEE-E20. EEE-E19의 방법에 있어서, 디커플링 툴은:

[0251] 명시된 주파수 아래의 독립적으로 디코딩된 신호를 추출하는 단계;

[0252] 명시된 주파수 위의 복합 신호를 추출하는 단계;

[0253] 복합 신호 그리고 복합 신호의 에너지 및 개개의 신호의 에너지에 관한 스케일 팩터로부터 명시된 주파수 위의 개개의 디커플링된 신호를 결정하는 단계; 및

[0254] 공동으로 디코딩된 신호를 획득하기 위해 각각의 독립적으로 디코딩된 신호를 개개의 디커플링된 신호와 결합하는 단계를 포함하는, 방법.

[0255] EEE-E21. EEE-E15의 방법에 있어서, 하나 이상의 오디오 신호 및 하나 이상의 추가적인 오디오 신호를 공동으로 디코딩하는 단계는 둘 초과의 오디오 신호를 추출하기 위해 공동-디코딩 툴을 적용하는 단계를 포함하는, 방법.

[0256] EEE-E22. EEE-E11의 방법에 있어서, 재생 디바이스와 연관된 하나 이상의 오디오 신호를 디코딩하는 단계는 오디오 신호가 코딩된 것과 동일한 도메인의 오디오 신호에 대역폭 확장을 적용하는 단계를 포함하는, 방법.

[0257] EEE-E23. EEE-E22의 방법에 있어서, 도메인은 변형 이산 코사인 변환(MDCT) 도메인인, 방법.

[0258] EEE-E24. EEE-E22 또는 EEE-E23의 방법에 있어서, 대역폭 확장은 적응형 잡음 추가를 포함하는, 방법.

[0259] EEE-E25. EEE-E1 내지 EEE-E24 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0260] EEE-E26. 실행될 때, 하나 이상의 디바이스가 EEE-E1 내지 EEE-E24 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0261] EEE-F1. 하나 이상의 마이크로폰을 갖는 디바이스에 의해 수행되는, 인코딩된 비트스트림을 생성하기 위한 방법으로서,

[0262] 하나 이상의 마이크로폰에 의해, 하나 이상의 오디오 신호를 캡처하는 단계;

[0263] 웨이크 워드의 존재를 결정하기 위해 캡처된 오디오 신호를 분석하는 단계;

[0264] 웨이크 워드의 존재를 검출시에:

[0265] 캡처된 오디오 신호에 대해 음성 인식 태스크가 수행되어야 함을 표시하기 위해 플래그를 설정하는 단계:

[0266] 캡처된 오디오 신호를 인코딩하는 단계;

[0267] 인코딩된 오디오 신호 및 플래그를 인코딩된 비트스트림으로 어셈블링하는 단계를 포함하는, 방법.

[0268] EEE-F2. EEE-F1의 방법에 있어서, 하나 이상의 마이크로폰은 모노 또는 공간 음장을 캡처하도록 구성되는, 방법.

[0269] EEE-F3. EEE-F2의 방법에 있어서, 공간 음장은 A-포맷 또는 B-포맷인, 방법.

[0270] EEE-F4. EEE-F1 내지 EEE-F3 중 어느 하나의 방법에 있어서, 캡처된 오디오 신호는 음성 인식 태스크를 수행하는 데에만 사용되도록 의도되는, 방법.

[0271] EEE-F5. EEE-F4의 방법에 있어서, 캡처된 오디오 신호는, 캡처된 오디오 신호가 디코딩될 때, 디코딩된 오디오 신호의 품질이 음성 인식 태스크를 수행하기에 충분하지만 사람의 청취를 위해서는 충분하지 않도록 인코딩되는, 방법.

[0272] EEE-F6. EEE-F4 또는 EEE-F5의 방법에 있어서, 캡처된 오디오 신호는 캡처된 오디오 신호를 인코딩하기 전에 대역 에너지, Mel-frequency Cepstral Coefficients, 또는 변형 이산 코사인 변환(MDCT) 스펙트럼 계수 중 하나 이상을 포함하는 표현으로 변환되는, 방법.

[0273] EEE-F7. EEE-F1 내지 EEE-F3 중 어느 하나의 방법에 있어서, 캡처된 오디오 신호는 사람의 청취를 위해 그리고 음성 인식 태스크를 수행하는 데 사용되도록 의도되는, 방법.

[0274] EEE-F8. EEE-F7의 방법에 있어서, 캡처된 오디오 신호는, 캡처된 오디오 신호가 디코딩될 때, 디코딩된 오디오 신호의 품질이 사람의 청취를 위해 충분하도록 인코딩되는, 방법.

[0275] EEE-F9. EEE-F7의 방법에 있어서, 캡처된 오디오 신호를 인코딩하는 단계는 캡처된 오디오 신호의 제1 인코딩된 표현 및 캡처된 오디오 신호의 제2 인코딩된 표현을 생성하는 단계를 포함하고, 제1 인코딩된 표현은, 캡처된 오디오 신호가 제1 인코딩된 표현으로부터 디코딩될 때, 디코딩된 오디오 신호의 품질이 사람의 청취를 위해 충분하도록 생성되고, 제2 인코딩된 표현은, 캡처된 오디오 신호가 제2 인코딩된 표현으로부터 디코딩될 때, 디코딩된 오디오 신호의 품질이 음성 인식 태스크를 수행하기에 충분하지만 사람의 청취를 위해서는 충분하지 않도록 생성되는, 방법.

[0276] EEE-F10. EEE-F9의 방법에 있어서, 캡처된 오디오 신호의 제2 인코딩된 표현을 생성하는 단계는, 캡처된 오디오 신호를 인코딩하기 전에, 캡처된 오디오 신호를 파라메트릭 표현, 조악한 파형 표현, 또는 대역 에너지, Mel-frequency Cepstral Coefficients, 또는 변형 이산 코사인 변환(MDCT) 스펙트럼 계수 중 하나 이상을 포함하는 표현 중 하나 이상으로 변환하는 단계를 포함하는, 방법.

[0277] EEE-F11. EEE-F9 또는 EEE-F10의 방법에 있어서, 인코딩된 오디오 신호를 비트스트림으로 어셈블링하는 단계는 제1 인코딩된 표현을 인코딩된 비트스트림의 제1 독립적인 블록 내로 삽입하는 단계, 및 제2 인코딩된 표현을 인코딩된 비트스트림의 제2 독립적인 블록 내로 삽입하는 단계를 포함하는, 방법.

[0278] EEE-F12. EEE-F9 또는 EEE-F10의 방법에 있어서, 제1 인코딩된 표현은 인코딩된 비트스트림의 제1 계층에 포함되고, 제2 인코딩된 표현은 인코딩된 비트스트림의 제2 계층에 포함되고, 제1 계층 및 제2 계층은 인코딩된 비트스트림의 단일 블록에 포함되는, 방법.

[0279] EEE-F13. EEE-F1 내지 EEE-F12 중 어느 하나의 방법에 있어서, 웨이크 워드의 존재가 검출되지 않을 때,

[0280] 캡처된 오디오 신호에 대해 음성 인식 태스크가 수행되지 않아야 함을 표시하기 위해 플래그를 설정하는 단계;

[0281] 캡처된 오디오 신호를 인코딩하는 단계;

[0282] 인코딩된 오디오 신호 및 플래그를 인코딩된 비트스트림으로 어셈블링하는 단계를 포함하는, 방법.

[0283] EEE-F14. 오디오 신호를 디코딩하기 위한 방법으로서,

[0284] 인코딩된 오디오 신호 및 음성 인식 태스크가 수행되어야 하는지를 표시하는 플래그를 포함하는 인코딩된 비트스트림을 수신하는 단계;

[0285] 디코딩된 오디오 신호를 획득하기 위해 인코딩된 오디오 신호를 디코딩하는 단계; 및

[0286] 플래그가 음성 인식 태스크가 수행되어야 함을 표시할 때, 디코딩된 오디오 신호에 대해 음성 인식 태스크를 수행하는 단계를 포함하는, 방법.

[0287] EEE-F15. EEE-F14의 방법에 있어서, 디코딩된 오디오 신호는 음성 인식 태스크를 수행하는 데에만 사용되도록

의도되는, 방법.

[0288] EEE-F16. EEE-F15의 방법에 있어서, 디코딩된 오디오 신호의 품질은 음성 인식 태스크를 수행하기에 충분하지만 사람의 청취를 위해서는 충분하지 않은, 방법.

[0289] EEE-F17. EEE-F15 또는 EEE-F16의 방법에 있어서, 디코딩된 오디오 신호는 캡처된 오디오 신호를 인코딩하기 전에 대역 에너지, Mel-frequency Cepstral Coefficients, 또는 변형 이산 코사인 변환(MDCT) 스펙트럼 계수 중 하나 이상을 포함하는 표현에 있는, 방법.

[0290] EEE-F18. EEE-F14의 방법에 있어서, 캡처된 오디오 신호는, 캡처된 오디오 신호가 디코딩될 때, 디코딩된 오디오 신호의 품질이 사람의 청취를 위해 충분하도록 인코딩되는, 방법.

[0291] EEE-F19. EEE-F18의 방법에 있어서, 인코딩된 오디오 신호는 하나 이상의 오디오 신호의 제1 인코딩된 표현 및 하나 이상의 오디오 신호의 제2 인코딩된 표현을 포함하는, 방법.

[0292] EEE-F20. EEE-F18의 방법에 있어서, 제1 표현으로부터 디코딩된 오디오 신호의 품질은 사람의 청취를 위해 충분하고, 제2 표현으로부터 디코딩된 오디오 신호의 품질은 음성 인식 태스크를 수행하기에 충분하지만 사람의 청취를 위해서는 충분하지 않은, 방법.

[0293] EEE-F21. EEE-F19 또는 EEE-F20의 방법에 있어서, 제1 표현은 인코딩된 비트스트림의 제1 독립적인 블록에 있고, 제2 표현은 인코딩된 비트스트림의 제2 독립적인 블록에 있는, 방법.

[0294] EEE-F22. EEE-F19 또는 EEE-F20의 방법에 있어서, 제1 표현은 인코딩된 비트스트림의 제1 계층에 있고, 제2 인코딩된 표현은 인코딩된 비트스트림의 제2 계층에 포함되고, 제1 계층 및 제2 계층은 인코딩된 비트스트림의 단일 블록에 포함되는, 방법.

[0295] EEE-F23. EEE-F18 내지 EEE-F22 중 어느 하나의 방법에 있어서, 인코딩된 오디오 신호를 디코딩하는 단계는 제2 표현만을 디코딩하는 단계, 및 제1 표현을 무시하는 단계를 포함하는, 방법.

[0296] EEE-F24. EEE-F18 내지 EEE-F23 중 어느 하나의 방법에 있어서, 제2 인코딩된 표현으로부터 디코딩된 오디오 신호는 파라메트릭 표현, 파형 표현, 또는 대역 에너지, Mel-frequency Cepstral Coefficients, 또는 변형 이산 코사인 변환(MDCT) 스펙트럼 계수 중 하나 이상을 포함하는 표현에 있는, 방법.

[0297] EEE-F25. EEE-F1 내지 EEE-F24 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0298] EEE-F26. 실행될 때, 하나 이상의 디바이스가 EEE-F1 내지 EEE-F24 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0299] EEE-G1. 하나 이상의 재생 디바이스로의 낮은 레이턴시 전송을 위한 몰입형 오디오 프로그램의 오디오 신호를 인코딩하기 위한 방법으로서,

[0300] 몰입형 오디오 프로그램의 복수의 시간-도메인 오디오 신호를 수신하는 단계;

[0301] 프레임 크기를 선택하는 단계;

[0302] 프레임 크기에 응답하여 시간-도메인 오디오 신호의 프레임을 추출하는 단계 - 시간-도메인 오디오 신호의 프레임은 시간-도메인 오디오 신호의 이전의 프레임과 중첩됨 -;

[0303] 오디오 신호를 중첩하는 프레임으로 세그먼트화하는 단계;

[0304] 시간-도메인 오디오 신호의 프레임을 주파수-도메인 신호로 변환하는 단계;

[0305] 주파수-도메인 신호를 코딩하는 단계;

[0306] 지각적으로 동기 부여된 양자화 툴을 사용하여 코딩된 주파수-도메인 신호를 양자화하는 단계;

[0307] 양자화되고 코딩된 주파수-도메인 신호를 프레임 내의 하나 이상의 독립적인 블록으로 어셈블링하는 단계; 및

[0308] 하나 이상의 독립적인 블록을 인코딩된 프레임으로 어셈블링하는 단계를 포함하는, 방법.

[0309] EEE-G2. EEE-G1의 방법에 있어서, 복수의 오디오 신호는 정의된 채널 구성을 갖는 채널 기반 신호를 포함하는, 방법.

[0310] EEE-G3. EEE-G2의 방법에 있어서, 채널 구성은 모노, 스테레오, 5.1, 5.1.2, 5.1.4, 7.1.2, 7.1.4, 9.1.6, 또

는 22.2 중 하나인, 방법.

[0311] EEE-G4. EEE-G1 내지 EEE-G3 중 어느 하나의 방법에 있어서, 복수의 오디오 신호는 하나 이상의 객체 기반 신호를 포함하는, 방법.

[0312] EEE-G5. EEE-G1 내지 EEE-G4 중 어느 하나의 방법에 있어서, 복수의 오디오 신호는 몰입형 오디오 프로그램의 장면 기반 표현을 포함하는, 방법.

[0313] EEE-G6. EEE-G1 내지 EEE-G5 중 어느 하나의 방법에 있어서, 선택된 프레임 크기는 128, 256, 512, 1024, 120, 240, 480, 또는 960 샘플 중 하나인, 방법.

[0314] EEE-G7. EEE-G1 내지 EEE-G6 중 어느 하나의 방법에 있어서, 시간-도메인 오디오 신호의 프레임과 시간-도메인 오디오 신호의 이전의 프레임 사이의 중첩은 50% 이하인, 방법.

[0315] EEE-G8. EEE-G1 내지 EEE-G7 중 어느 하나의 방법에 있어서, 변환은 변형 이산 코사인 변환(MDCT)인, 방법.

[0316] EEE-G9. EEE-G1 내지 EEE-G8 중 어느 하나의 방법에 있어서, 복수의 오디오 신호 중 둘 이상이 공동으로 코딩되는, 방법.

[0317] EEE-G10. EEE-G1 내지 EEE-G9 중 어느 하나의 방법에 있어서, 각각의 독립적인 블록은 하나 이상의 재생 디바이스에 대한 인코딩된 신호를 포함하는, 방법.

[0318] EEE-G11. EEE-G1 내지 EEE-G10 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 둘 이상의 재생 디바이스에 대한 인코딩된 신호를 포함하고, 인코딩된 신호는 공동으로 코딩된 오디오 신호를 포함하는, 방법.

[0319] EEE-G12. EEE-G1 내지 EEE-G11 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 재생 디바이스의 상이한 드라이버로부터의 재생을 위해 의도된 상이한 대역폭을 커버하는 복수의 인코딩된 신호를 포함하는, 방법.

[0320] EEE-G13. EEE-G1 내지 EEE-G12 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 재생 디바이스에 의해 수행되는 에코-관리에 사용하기 위한 인코딩된 에코-참조 신호를 포함하는, 방법.

[0321] EEE-G14. EEE-G1 내지 EEE-G13 중 어느 하나의 방법에 있어서, 양자화된 주파수-도메인 신호를 코딩하는 단계는 다음의 툴: 시간 잡음 성형(TNS), 공동-채널 코딩, 신호에 걸친 스케일 팩터의 공유, 고주파수 재구성을 위한 제어 파라미터의 결정, 및 잡음 치환을 위한 제어 파라미터의 결정 중 하나 이상을 적용하는 단계를 포함하는, 방법.

[0322] EEE-G15. EEE-G1 내지 EEE-G14 중 어느 하나의 방법에 있어서, 하나 이상의 독립적인 블록은 재생 디바이스의 지연, 이득, 및 동등화 중 하나 이상을 제어하기 위한 파라미터를 포함하는, 방법.

[0323] EEE-G16. 인코딩된 신호로부터 몰입형 오디오 프로그램의 오디오 신호를 디코딩하기 위한 낮은 레이턴시 방법으로서,

[0324] 하나 이상의 독립적인 블록을 포함하는 인코딩된 프레임을 수신하는 단계;

[0325] 하나 이상의 독립적인 블록으로부터, 양자화되고 코딩된 주파수-도메인 신호를 추출하는 단계;

[0326] 양자화되고 코딩된 주파수-도메인 신호를 역 양자화하는(dequantizing) 단계;

[0327] 역 양자화된 주파수-도메인 신호를 디코딩하는 단계;

[0328] 시간-도메인 신호를 획득하기 위해 디코딩된 주파수-도메인 신호를 역 변환하는 단계; 및

[0329] 몰입형 오디오 프로그램의 복수의 오디오 신호를 제공하기 위해 이전의 프레임으로부터의 시간-도메인 신호에 시간-도메인 신호를 중첩하고 추가하는 단계를 포함하는, 방법.

[0330] EEE-G17. EEE-G16의 방법에 있어서, 복수의 오디오 신호는 정의된 채널 구성을 갖는 채널 기반 신호를 포함하는, 방법.

[0331] EEE-G18. EEE-G17의 방법에 있어서, 채널 구성은 모노, 스테레오, 5.1, 5.1.2, 5.1.4, 7.1.2, 7.1.4, 9.1.6, 또는 22.2 중 하나인, 방법.

[0332] EEE-G19. EEE-G16 내지 EEE-G18 중 어느 하나의 방법에 있어서, 복수의 오디오 신호는 하나 이상의 객체 기반

신호를 포함하는, 방법.

[0333] EEE-G20. EEE-G16 내지 EEE-G19 중 어느 하나의 방법에 있어서, 복수의 오디오 신호는 몰입형 오디오 프로그램의 장면 기반 표현을 포함하는, 방법.

[0334] EEE-G21. EEE-G16 내지 EEE-G20 중 어느 하나의 방법에 있어서, 시간-도메인 샘플의 프레임은 128, 256, 512, 1024, 120, 240, 480, 또는 960 샘플 중 하나를 포함하는, 방법.

[0335] EEE-G22. EEE-G16 내지 EEE-G21 중 어느 하나의 방법에 있어서, 이전의 프레임과의 중첩이 50% 이하인, 방법.

[0336] EEE-G23. EEE-G16 내지 EEE-G22 중 어느 하나의 방법에 있어서, 역 변환은 역 변형 이산 코사인 변환(IMDCT)인, 방법.

[0337] EEE-G24. EEE-G16 내지 EEE-G23 중 어느 하나의 방법에 있어서, 각각의 독립적인 블록은 하나 이상의 재생 디바이스에 대한 양자화되고 코딩된 주파수-도메인 신호를 포함하는, 방법.

[0338] EEE-G25. EEE-G16 내지 EEE-G24 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 둘 이상의 재생 디바이스에 대한 양자화되고 코딩된 주파수-도메인 신호를 포함하고, 양자화되고 코딩된 주파수-도메인 신호는 공동으로 코딩된 오디오 신호인, 방법.

[0339] EEE-G26. EEE-G16 내지 EEE-G25 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 재생 디바이스의 상이한 드라이버로부터의 재생을 위해 의도된 상이한 대역폭을 커버하는 복수의 양자화되고 코딩된 주파수-도메인 신호를 포함하는, 방법.

[0340] EEE-G27. EEE-G16 내지 EEE-G26 중 어느 하나의 방법에 있어서, 적어도 하나의 독립적인 블록은 재생 디바이스에 의해 수행되는 에코-관리에 사용하기 위한 인코딩된 에코-참조 신호를 포함하는, 방법.

[0341] EEE-G28. EEE-G16 내지 EEE-G27 중 어느 하나의 방법에 있어서, 역 양자화된 주파수-도메인 신호를 디코딩하는 단계는 다음의 디코딩 툴: 시간 잡음 성형(TNS), 공동-채널 디코딩, 신호에 걸친 스케일 팩터의 공유, 고주파 재구성, 및 잡음 치환 중 하나 이상을 적용하는 단계를 포함하는, 방법.

[0342] EEE-G29. EEE-G16 내지 EEE-G28 중 어느 하나의 방법에 있어서, 하나 이상의 독립적인 블록은 재생 디바이스의 지연, 이득, 및 동등화 중 하나 이상을 제어하기 위한 파라미터를 포함하는, 방법.

[0343] EEE-G30. EEE-G16 내지 EEE-G29 중 어느 하나의 방법에 있어서, 방법은 재생 디바이스에 의해 수행되고, 하나 이상의 독립적인 블록으로부터 양자화되고 코딩된 신호를 추출하는 단계는 재생 디바이스에 의한 재생을 위해 양자화되고 코딩된 주파수-도메인 신호를 포함하는 블록만을 선택하는 단계, 및 다른 재생 디바이스에 의한 재생을 위해 양자화되고 코딩된 주파수-도메인 신호를 포함하는 독립적인 블록을 무시하는 단계를 포함하는, 방법.

[0344] EEE-G31. EEE-G1 내지 EEE-G30 중 어느 하나의 방법을 수행하도록 구성된 장치.

[0345] EEE-G32. 실행될 때, 하나 이상의 디바이스가 EEE-G1 내지 EEE-G30 중 어느 하나의 방법을 수행하는 것을 야기하는 명령의 시퀀스를 포함하는, 비일시적 컴퓨터 판독가능 저장 매체.

[0346] 아래의 청구범위 및 본원의 설명에서, 포함하는(comprising, which comprises) 또는 -로 구성된(comprised of) 이란 용어 중 임의의 것은 적어도 뒤따르는 요소/특징을 적어도 포함하지만 다른 것 또한 제외하지 않는 것을 의미하는 개방형 용어이다. 따라서, 청구범위에서 사용될 때 포함하는 이란 용어는 이후에 나열된 수단 또는 요소 또는 단계로 제한하는 것으로 해석되어서는 안 된다. 예를 들어, A 및 B를 포함하는 디바이스의 표현의 범주는 요소 A 및 B로만 구성된 디바이스로 제한되어서는 안 된다. 본원에서 사용되는 바에 따르면 포함하는(including, which includes, that includes) 이란 용어 중 임의의 것은 또한 용어 뒤에 오는 요소/특징을 적어도 포함하지만 다른 것을 배제하지 않는 것을 의미하는 개방형 용어이다. 따라서 포함하는(including)은 포함하는(comprising)과 동의어이며 이를 의미한다.

[0347] 본 발명의 예의 위의 설명에서, 다양한 특징은 본 개시를 간소화하고 다양한 발명 양태 중 하나 이상의 이해를 돕기 위한 목적으로 때때로 단일의 예, 도면 또는 그의 설명으로 함께 그룹화되는 것으로 이해되어야 한다. 하지만, 본 개시의 이 방법은 각각의 청구항에 명시적으로 인용된 것보다 더 많은 특징이 요구된다는 의도를 반영하는 것으로 해석되어서는 안 된다. 오히려, 다음의 청구범위가 반영하는 바와 같이, 발명 양태는 단일의 전술한 개시된 예의 모든 특징보다 적은 부분에 있다. 따라서, 상세한 설명에 후속하는 청구범위는 본원에서 이 상

세한 설명에 명시적으로 통합되며, 각각의 청구항은 본 발명의 별개의 예로서 그 자체로 존재한다.

[0348] 또한, 본원에 설명된 일부 예는 다른 예에 포함된 다른 특징이 아닌 어떠한 특징을 포함하지만, 상이한 예의 특징의 조합은 포괄되는 것을 의미하며, 통상의 기술자에 의해 이해되는 바와 같이, 상이한 예를 형성한다. 예를 들어, 다음의 청구범위에서, 청구된 예 중 임의의 것이 임의의 조합으로 사용될 수 있다.

[0349] 또한, 예 중 일부는 본원에서, 컴퓨터 시스템의 프로세서에 의해 또는 기능을 수행하는 다른 수단에 의해 구현될 수 있는 방법 또는 방법의 요소의 조합으로 설명된다. 따라서, 이러한 방법 또는 방법의 요소를 수행하는 데 필요한 명령어를 갖는 프로세서는 방법 또는 방법의 요소를 수행하기 위한 수단을 형성한다. 또한, 장치의 본원에서 설명된 요소는 요소에 의해 수행되는 기능을 실행하기 위한 수단의 예이다.

[0350] 또한, 본원에 설명된 예 중 일부는 유선 및/또는 무선 시스템과 같은 분배 및/또는 전송 시스템에서 가능하게는 구현되는 것으로 해석되어야 하는 연결된 솔루션을 개시한다. 예를 들어, 5G 뿐만 아니라 3G, 4G와 같은 임의의 전기, 광학, 및/또는 모바일 시스템의 사용에 의한 것이다.

[0351] 따라서, 본 발명의 특정 예가 설명되었지만, 통상의 기술자는 다른 및 추가적인 수정이 이루어질 수 있다는 것을 인식할 것이고, 이러한 모든 변경 및 수정을 청구하는 것이 의도된다. 예를 들어, 위에서 주어진 임의의 식은 단지 사용될 수 있는 절차를 표현하는 것이다. 블록도로부터 기능이 추가되거나 삭제될 수 있고, 기능 블록 사이에 동작이 교환될 수 있다. 단계는 설명된 방법에 추가되거나 삭제될 수 있다.

[0352] 앞서 개시된 시스템, 디바이스, 및 방법은 소프트웨어, 펌웨어, 하드웨어 또는 이들의 조합으로 구현될 수 있다. 예를 들어, 본 출원의 양태는, 적어도 부분적으로, 디바이스, 하나 초과의 디바이스를 포함하는 시스템, 방법, 컴퓨터 프로그램 제품 등으로 구현될 수 있다.

[0353] 하드웨어 구현에서, 위의 설명에서 참조된 기능적 유닛들 사이의 작업의 분할이 물리적인 유닛으로의 분할에 반드시 대응하는 것은 아니고; 대조적으로, 하나의 물리적인 구성요소는 다수의 기능을 가질 수 있고, 하나의 작업은 여러 물리적인 구성요소에 의해 협력하여 수행될 수 있다.

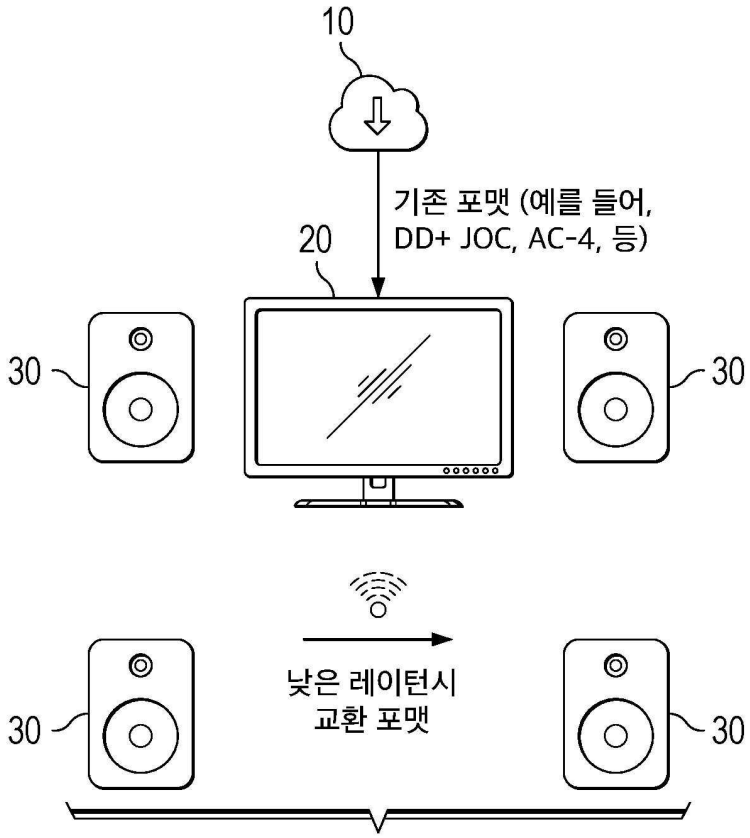
[0354] 특정한 구성요소 또는 모든 구성요소는 디지털 신호 프로세서 또는 마이크로프로세서에 의해 실행되는 소프트웨어로 구현되거나 하드웨어 또는 주문형 집적 회로로 구현될 수 있다. 이러한 소프트웨어는 컴퓨터 저장 매체(또는 비일시적 매체) 및 통신 매체(또는 일시적 매체)를 포함할 수 있는, 컴퓨터 판독가능 매체 상에 배포될 수 있다.

[0355] 통상의 기술자에게 잘 알려진 바와 같이, 컴퓨터 저장 매체라는 용어는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 다른 데이터와 같이, 정보의 저장을 위한 임의의 방법 또는 기술에서 구현된 휘발성 및 비휘발성, 이동식 및 비-이동식 매체 모두를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 다른 메모리 기술, CD-ROM, 디지털 다목적 디스크(digital versatile disks, DVD) 또는 다른 광학 디스크 저장소, 자기 카세트, 자기 테이프, 자기 디스크 저장소 또는 다른 자기 저장 디바이스 또는 원하는 정보를 저장하는 데 사용될 수 있고 컴퓨터에 의해 액세스될 수 있는 임의의 다른 매체를 포함하지만 이에 제한되지 않는다.

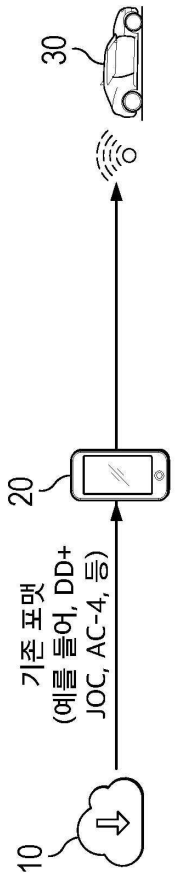
[0356] 추가로, 통신 매체는 통상적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파 또는 기타 전송 메커니즘과 같은 변조된 데이터 신호의 다른 데이터를 구현하고, 임의의 정보 전달 매체를 포함한다는 것이 통상의 기술자에게 잘 공지된다.

도면

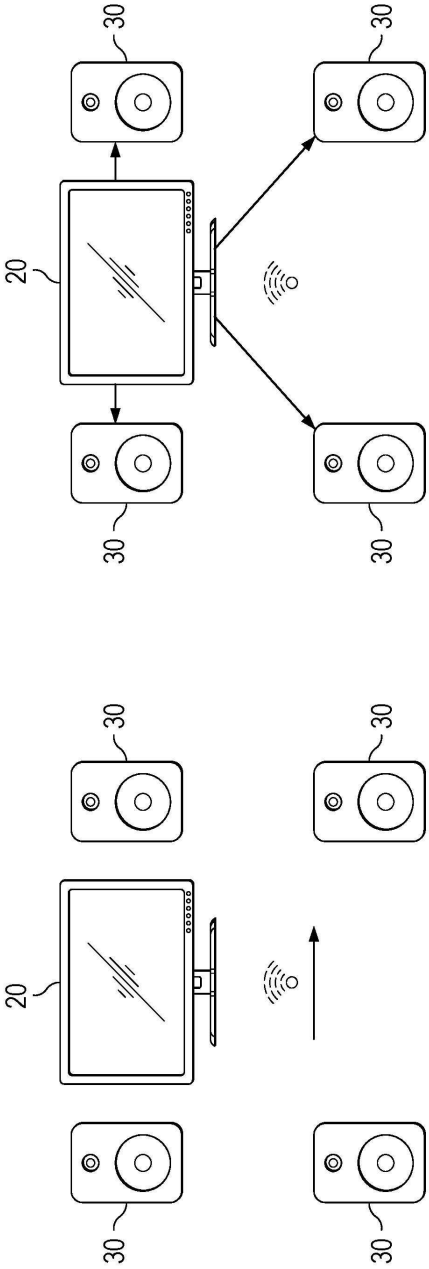
도면1



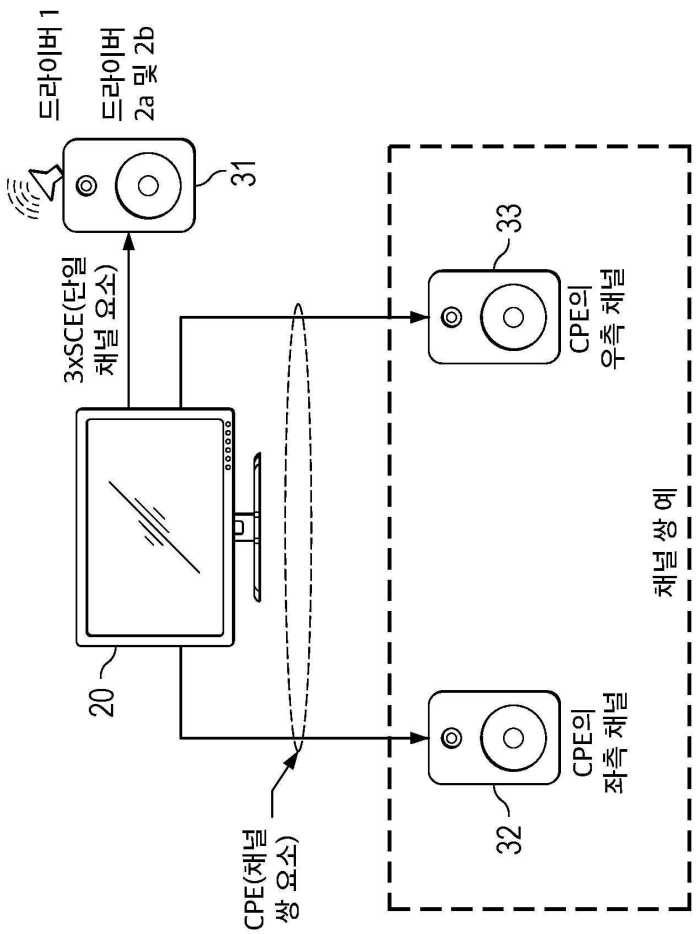
도면2



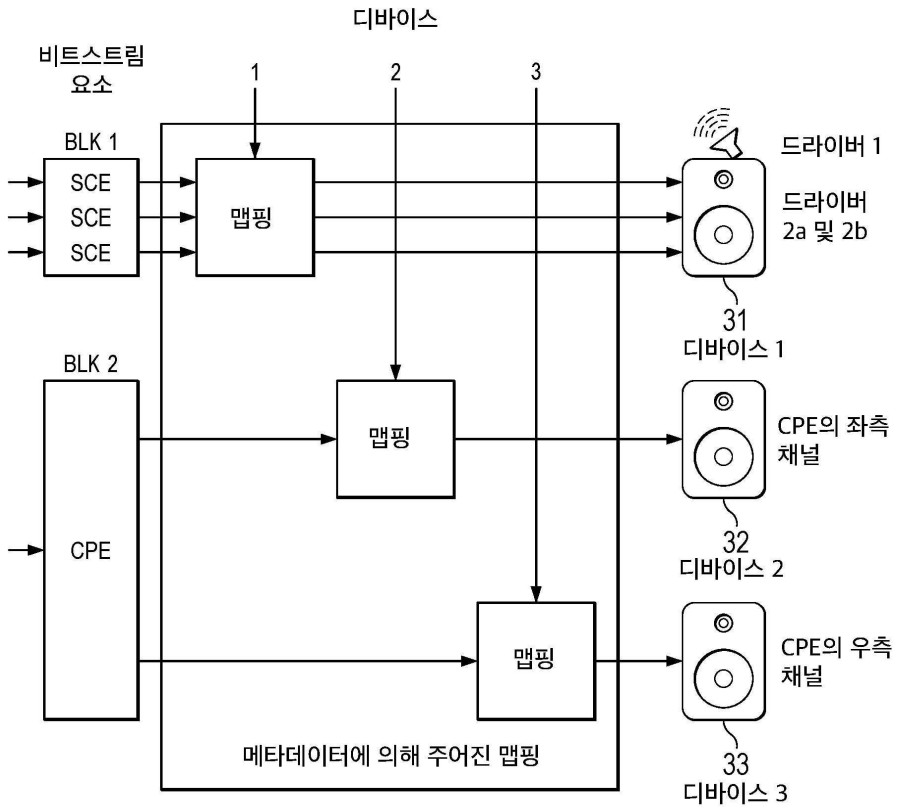
도면3



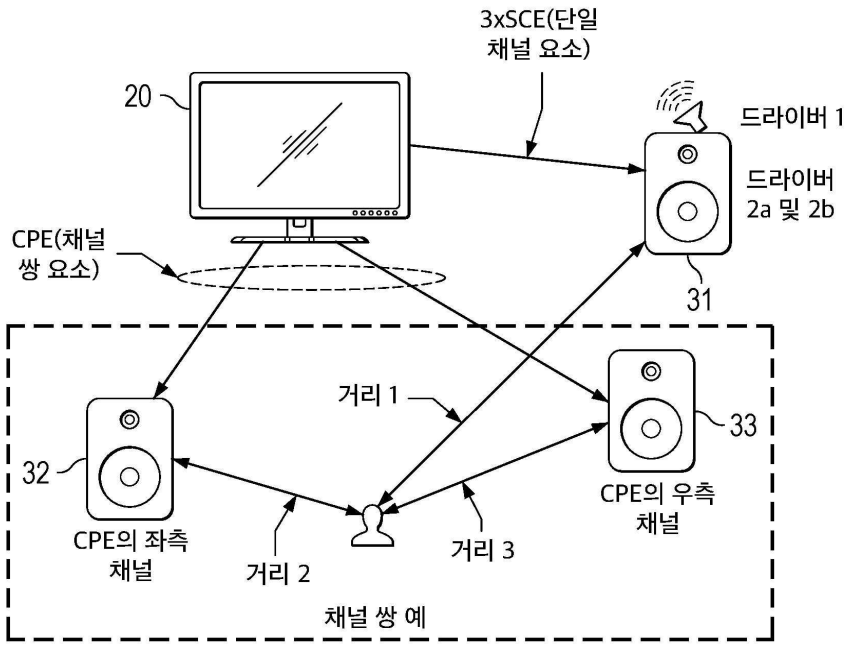
도면4


드라이버 1
드라이버
2a 및 2b
31
3xSCE(단일
채널 요소)
20
CPE(채널
쌍 요소)
33
CPE의
우측 채널
32
CPE의
좌측 채널
채널 쌍 예

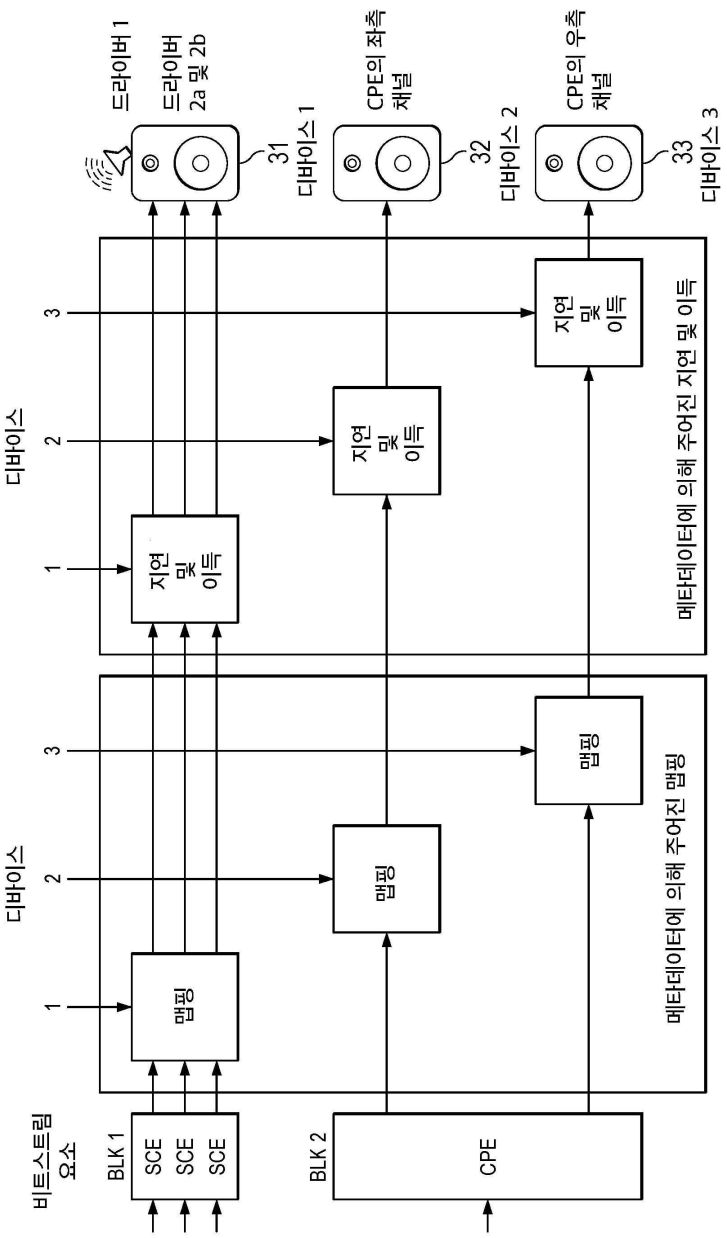
도면5


디바이스
비트스트림
요소
1
2
3
BLK 1
SCE
SCE
SCE
맵핑
드라이버 1
드라이버
2a 및 2b
31
디바이스 1
BLK 2
CPE
맵핑
CPE의 좌측
채널
32
디바이스 2
맵핑
CPE의 우측
채널
33
디바이스 3
메타데이터에 의해 주어진 맵핑

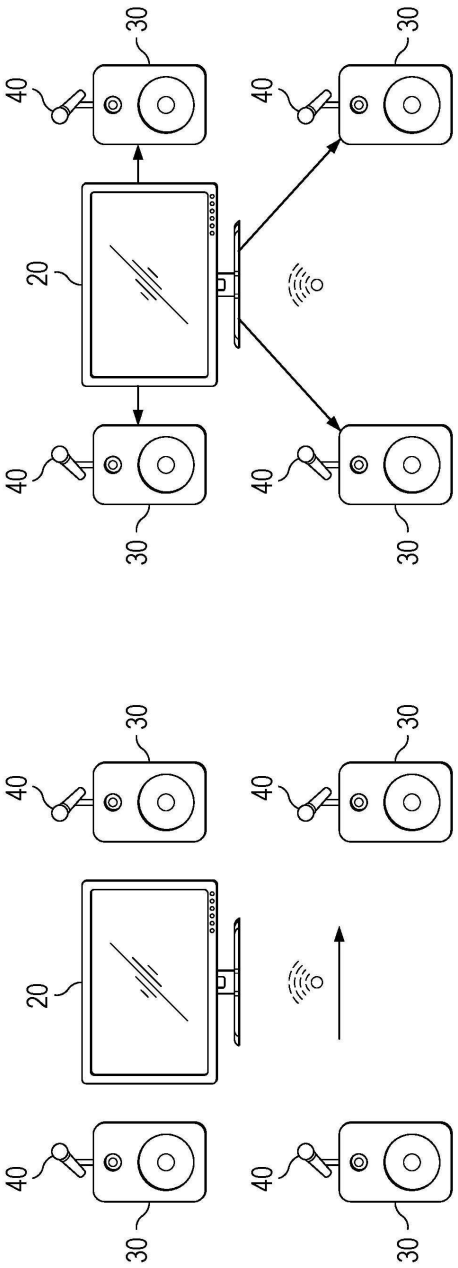
도면6


3xSCE(단일
채널 요소)
20
드라이버 1
드라이버
2a 및 2b
CPE(채널
쌍 요소)
31
거리 1
33
32
CPE의 우측
채널
CPE의 좌측
채널
거리 2
거리 3
채널 쌍 예

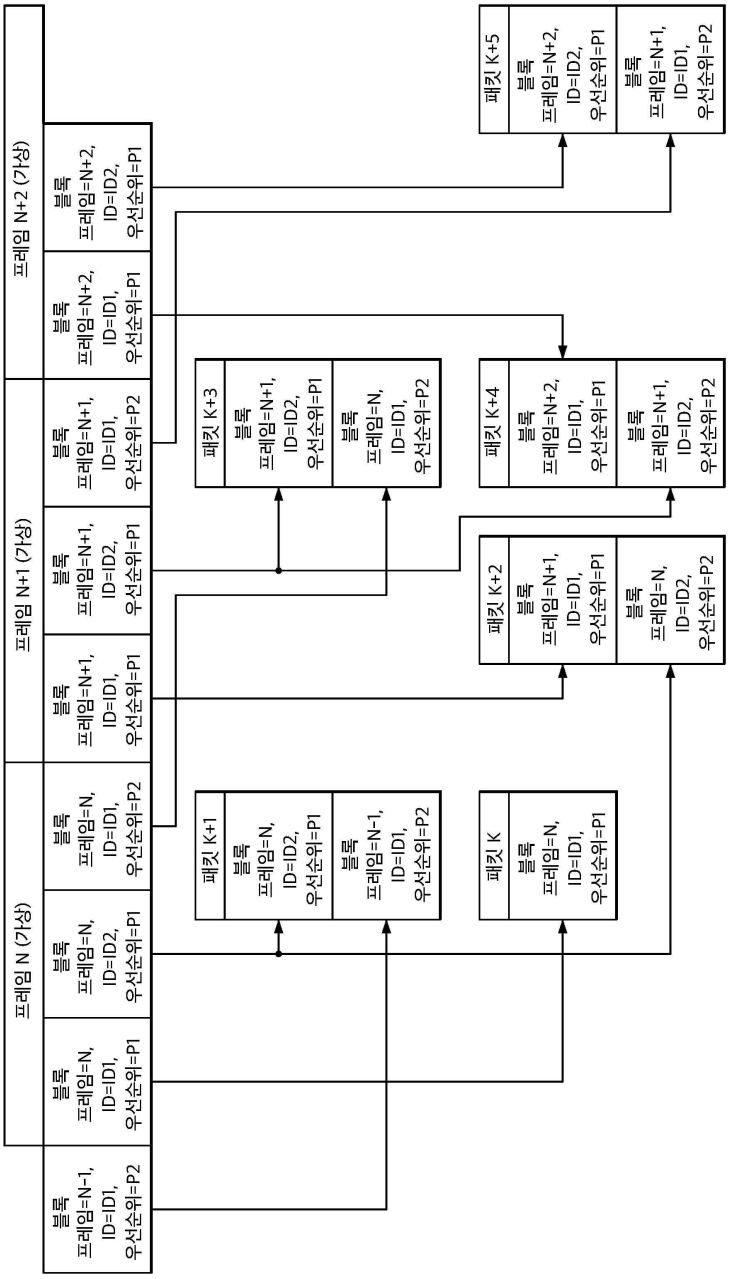
도면7


비트스트림 요소
디바이스
1
2
3
BLK 1
SCE
SCE
SCE
맵핑
맵핑
지연 및 이득
BLK 2
CPE
맵핑
지연 및 이득
메타데이터에 의해 주어진 맵핑
메타데이터에 의해 주어진 지연 및 이득
디바이스
1
2
3
지연 및 이득
지연 및 이득
지연 및 이득
드라이버 1
드라이버 2a 및 2b
31
디바이스 1
CPE의 좌측 채널
32
디바이스 2
CPE의 우측 채널
33
디바이스 3

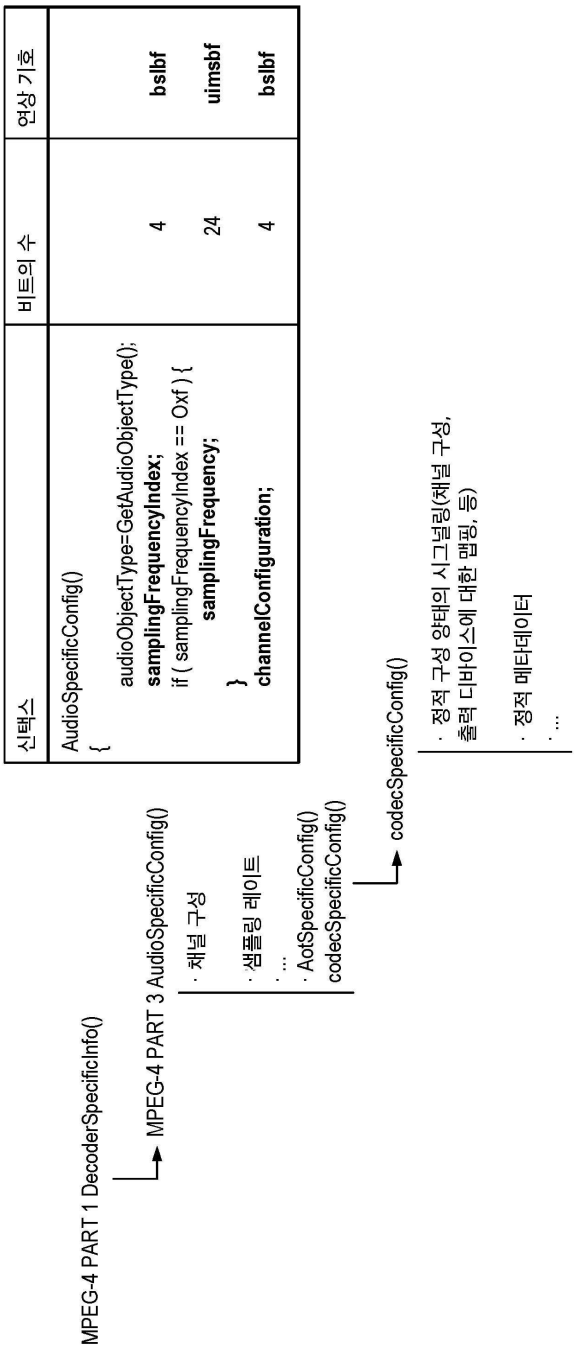
도면8



도면9


프레임 N (가상)
프레임 N+1 (가상)
프레임 N+2 (가상)
블록 프레임=N-1, ID=ID1, 우선순위=P2
블록 프레임=N, ID=ID1, 우선순위=P1
블록 프레임=N, ID=ID2, 우선순위=P1
블록 프레임=N, ID=ID1, 우선순위=P2
블록 프레임=N+1, ID=ID1, 우선순위=P1
블록 프레임=N+1, ID=ID2, 우선순위=P1
블록 프레임=N+1, ID=ID1, 우선순위=P2
블록 프레임=N+2, ID=ID1, 우선순위=P1
블록 프레임=N+2, ID=ID2, 우선순위=P1
패킷 K
블록 프레임=N, ID=ID1, 우선순위=P1
패킷 K+1
블록 프레임=N, ID=ID2, 우선순위=P1
블록 프레임=N-1, ID=ID1, 우선순위=P2
패킷 K+2
블록 프레임=N+1, ID=ID1, 우선순위=P1
블록 프레임=N, ID=ID2, 우선순위=P2
패킷 K+3
블록 프레임=N+1, ID=ID2, 우선순위=P1
블록 프레임=N, ID=ID1, 우선순위=P2
패킷 K+4
블록 프레임=N+2, ID=ID1, 우선순위=P1
블록 프레임=N+1, ID=ID2, 우선순위=P2
패킷 K+5
블록 프레임=N+2, ID=ID2, 우선순위=P1
블록 프레임=N+1, ID=ID1, 우선순위=P2

도면10


신택스
비트의 수
연상 기호
AudioSpecificConfig()
{
audioObjectType=GetAudioObjectType();
samplingFrequencyIndex;
4
bslbf
if ( samplingFrequencyIndex == 0xf ) {
samplingFrequency;
24
uimsbf
}
channelConfiguration;
4
bslbf
MPEG-4 PART 1 DecoderSpecificInfo()
MPEG-4 PART 3 AudioSpecificConfig()
· 채널 구성
· 샘플링 레이트
· ...
· AotSpecificConfig()
codecSpecificConfig()
codecSpecificConfig()
· 정적 구성 양태의 시그널링(채널 구성, 출력 디바이스에 대한 맵핑, 등)
· 정적 메타데이터
· ...

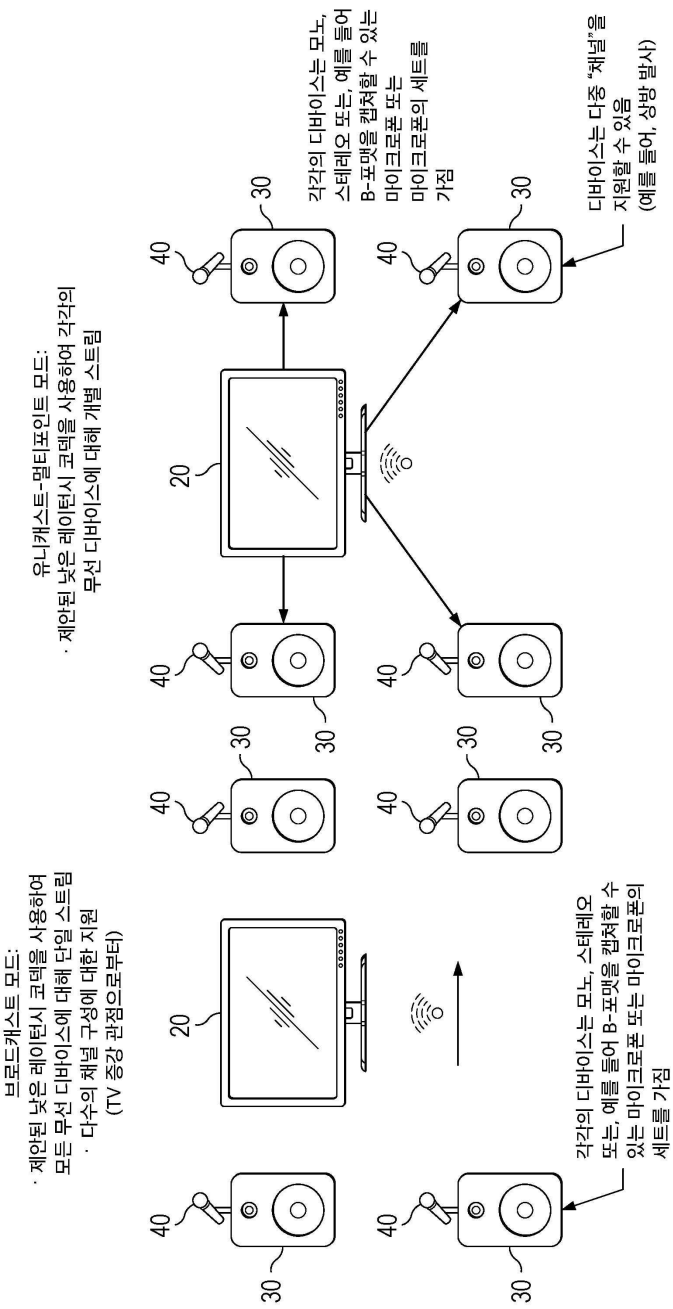
도면11

raw_data_block_codec(configData)

- MPEG-4의 기존 로우 데이터 블록과 유사
- 주어진 디바이스에 대해 필요하지 않은 데이터 블록의 부분을 스킵하는 것을 가능하게 하기 위해 추가적인 정보를 가짐
- 동적 메타데이터

| 신택스 | 비트의 수 | 연상 기호 |
|---|---|---|
| raw_data_block()<br>{<br>while( (id= **id_syn_ele**) != ID_END ){<br>switch (id) {<br>case ID_SCE: single_channel_element();<br>break;<br>case ID_CPE: channel_pair_element();<br>break;<br>case ID_CCE: coupling_channel_element();<br>break;<br>case ID_LFE: life_channel_element();<br>break;<br>case ID_DSE: data_stream_element();<br>break;<br>case ID_PCE: program_config_element();<br>break;<br>case ID_FIL: fill_element(); | 3 | **uimsbf** |

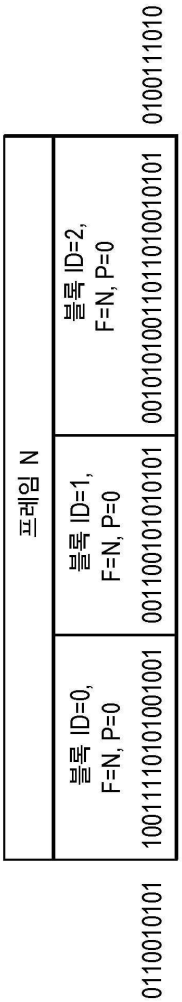
도면12


브로드캐스트 모드:
· 제안된 낮은 레이턴시 코덱을 사용하여 모든 무선 디바이스에 대해 단일 스트림
· 다수의 채널 구성에 대한 지원 (TV 증강 관점으로부터)
유니캐스트-멀티포인트 모드:
· 제안된 낮은 레이턴시 코덱을 사용하여 각각의 무선 디바이스에 대해 개별 스트림
각각의 디바이스는 모노, 스테레오 또는, 예를 들어 B-포맷을 캡처할 수 있는 마이크로폰 또는 마이크로폰의 세트를 가짐
디바이스는 다중 "채널"을 지원할 수 있음 (예를 들어, 상방 발사)
각각의 디바이스는 모노, 스테레오 또는, 예를 들어 B-포맷을 캡처할 수 있는 마이크로폰 또는 마이크로폰의 세트를 가짐
20
30
40

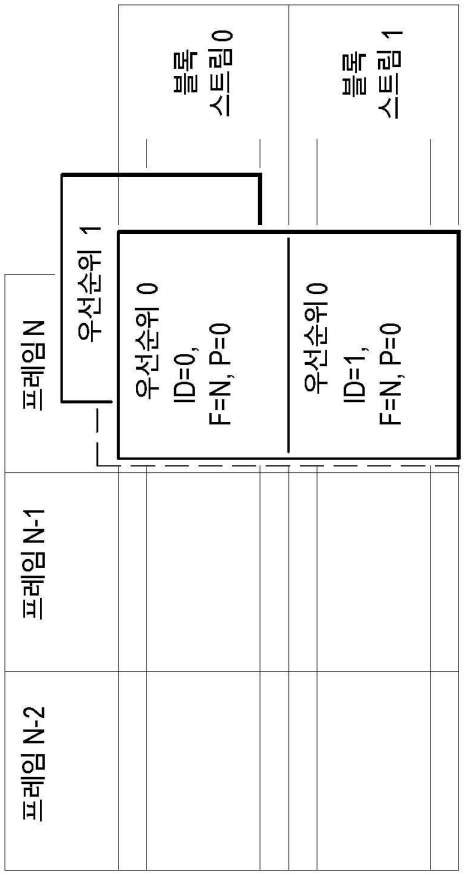
도면13

| 프레임 N-2 | 프레임 N-1 | 프레임 N | |
|---|---|---|---|
| | | | |
| | | ID=0,<br>F=N, P=0 | 블록<br>스트림 0 |
| | | | |
| | | | |
| | | ID=1,<br>F=N, P=0 | 블록<br>스트림 1 |
| | | | |
| | | | |
| | | ID=2,<br>F=N, P=0 | 블록<br>스트림 2 |
| | | | |

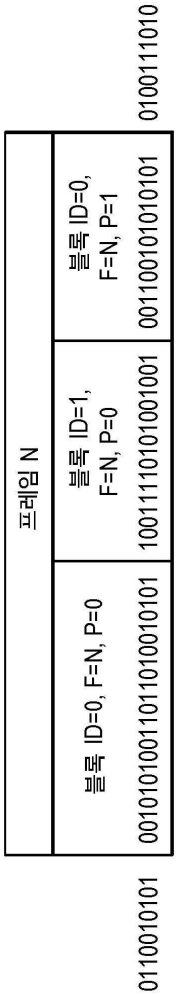
도면14



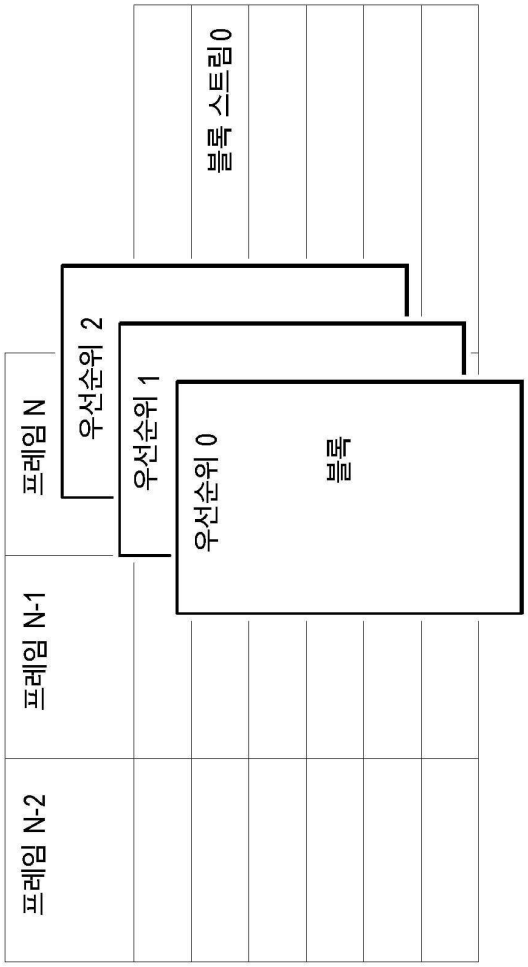
도면15


프레임 N
프레임 N-1
프레임 N-2
우선순위 1
우선순위 0
ID=0,
F=N, P=0
우선순위 0
ID=1,
F=N, P=0
블록
스트림 0
블록
스트림 1

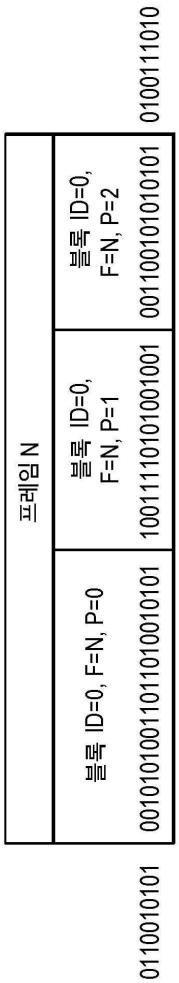
도면16


프레임 N
블록 ID=0, F=N, P=0
0010101001101101010010101
블록 ID=1, F=N, P=0
1001111010101001001
블록 ID=0, F=N, P=1
0011001010101010101
0110010101
0100111010

도면17



도면18


프레임 N
블록 ID=0, F=N, P=0
0010101001101101010010101
블록 ID=0, F=N, P=1
1001111010100 1001
블록 ID=0, F=N, P=2
0011001010101 01
0110010101
0100111010