



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0060502  
(43) 공개일자 2023년05월04일

- |  |  |
|--|--|
| (51) 국제특허분류(Int. Cl.)<br>G10L 19/008 (2014.01) G10L 21/0388 (2013.01)<br>H04S 7/00 (2006.01) | (71) 출원인<br>소니그룹주식회사<br>일본국 도쿄도 미나토쿠 코난 1-7-1                  |
| (52) CPC특허분류<br>G10L 19/008 (2020.08)<br>G10L 21/0388 (2013.01)                              | (72) 발명자<br>혼마 히로유키<br>일본 1080075 도쿄도 미나토쿠 코난 1-7-1 소니그룹주식회사 내 |
| (21) 출원번호 10-2023-7005227  | 치넨 도루<br>일본 1080075 도쿄도 미나토쿠 코난 1-7-1 소니그룹주식회사 내               |
| (22) 출원일자(국제) 2021년08월20일<br>심사청구일자 없음   | 고노 아키후미<br>일본 1080075 도쿄도 미나토쿠 코난 1-7-1 소니그룹주식회사 내             |
| (85) 번역문제출일자 2023년02월14일   | (74) 대리인<br>양영준, 이중희   |
| (86) 국제출원번호 PCT/JP2021/030599  |  |
| (87) 국제공개번호 WO 2022/050087<br>국제공개일자 2022년03월10일   |  |
| (30) 우선권주장<br>JP-P-2020-148234 2020년09월03일 일본(JP)  |  |

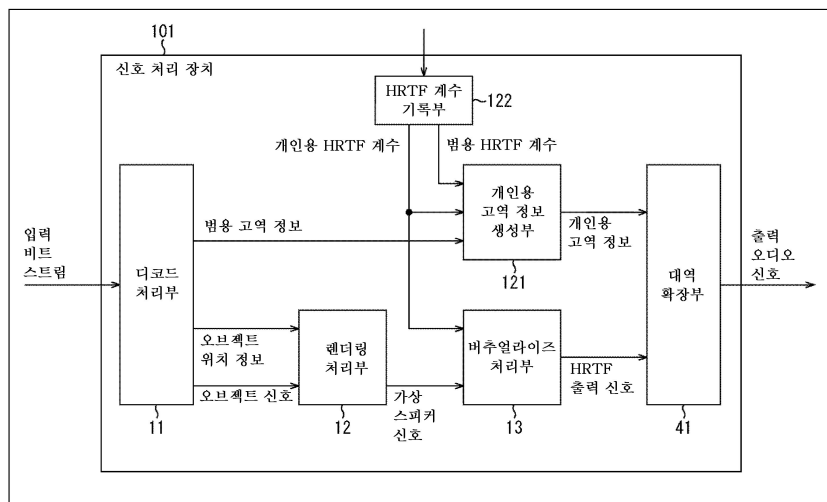
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 신호 처리 장치 및 방법, 학습 장치 및 방법, 그리고 프로그램

(57) 요약

본 기술은, 저비용의 장치에서도 고품질의 오디오 재생을 행할 수 있도록 하는 신호 처리 장치 및 방법, 학습 장치 및 방법, 그리고 프로그램에 관한 것이다. 신호 처리 장치는, 입력 비트 스트림을 제1 오디오 신호와, 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하는 디코딩 처리부와, 제1 오디오 신호 및 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 대역 확장부를 구비한다. 본 기술은 스마트폰에 적용할 수 있다.

대표도



(52) CPC특허분류  
*H04S 7/00* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하는 디코딩 처리부와,

상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 대역 확장부를 구비하는, 신호 처리 장치.

#### 청구항 2

제1항에 있어서, 상기 제1 고역 정보에 기초하여 상기 제2 고역 정보를 생성하는 고역 정보 생성부를 더 구비하는, 신호 처리 장치.

#### 청구항 3

제2항에 있어서, 상기 제1 고역 정보는, 제1 계수를 사용하여 상기 신호 처리를 행함으로써 얻어지는 상기 제2 오디오 신호의 대역 확장을 위한 고역 정보이고,

상기 제2 고역 정보는, 제2 계수를 사용하여 상기 신호 처리를 행함으로써 얻어지는 상기 제2 오디오 신호의 대역 확장을 위한 고역 정보이고,

상기 대역 확장부는, 상기 제1 오디오 신호, 상기 메타데이터 및 상기 제2 계수에 기초하여 상기 신호 처리를 행함으로써 얻어진 상기 제2 오디오 신호와, 상기 제2 고역 정보에 기초하여 상기 대역 확장 처리를 행하는, 신호 처리 장치.

#### 청구항 4

제3항에 있어서, 상기 고역 정보 생성부는, 상기 제1 고역 정보, 상기 제1 계수 및 상기 제2 계수에 기초하여, 상기 제2 고역 정보를 생성하는, 신호 처리 장치.

#### 청구항 5

제3항에 있어서, 상기 고역 정보 생성부는, 미리 기계 학습에 의해 생성된 계수와, 상기 제1 고역 정보, 상기 제1 계수 및 상기 제2 계수에 기초하는 연산을 행함으로써, 상기 제2 고역 정보를 생성하는, 신호 처리 장치.

#### 청구항 6

제5항에 있어서, 상기 연산은 뉴럴 네트워크에 의한 연산인, 신호 처리 장치.

#### 청구항 7

제3항에 있어서, 상기 제1 계수는 범용의 계수이고, 상기 제2 계수는 유저마다의 계수인, 신호 처리 장치.

#### 청구항 8

제7항에 있어서, 상기 제1 계수 및 상기 제2 계수는 HRTF 계수인, 신호 처리 장치.

#### 청구항 9

제1항에 있어서, 상기 신호 처리를 행하여 상기 제2 오디오 신호를 생성하는 신호 처리부를 더 구비하는, 신호 처리 장치.

#### 청구항 10

제9항에 있어서, 상기 신호 처리는, 베티컬라이즈 처리 또는 렌더링 처리를 포함하는 처리인, 신호 처리 장치.

**청구항 11**

제1항에 있어서, 상기 제1 오디오 신호는, 오디오 오브젝트의 오브젝트 신호, 또는 채널 베이스의 오디오 신호인, 신호 처리 장치.

**청구항 12**

신호 처리 장치가,

입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하고,

상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는, 신호 처리 방법.

**청구항 13**

입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하고,

상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 스텝을 포함하는 처리를 컴퓨터에 실행시키는, 프로그램.

**청구항 14**

제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하는 제1 고역 정보 계산부와,

상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하는 제2 고역 정보 계산부와,

상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는 고역 정보 학습부를 구비하는, 학습 장치.

**청구항 15**

제14항에 있어서, 상기 계수 데이터는 뉴럴 네트워크를 구성하는 계수인, 학습 장치.

**청구항 16**

제14항에 있어서, 상기 제1 계수는 범용의 계수이고, 상기 제2 계수는 유저마다의 계수인, 학습 장치.

**청구항 17**

제16항에 있어서, 상기 신호 처리는, 버추얼라이즈 처리 또는 렌더링 처리를 포함하는 처리이고,

상기 제1 계수 및 상기 제2 계수는 HRTF 계수인, 학습 장치.

**청구항 18**

제14항에 있어서, 상기 제1 오디오 신호는, 오디오 오브젝트의 오브젝트 신호, 또는 채널 베이스의 오디오 신호인, 학습 장치.

**청구항 19**

학습 장치가,

제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하고,

상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하고,

상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교차 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는, 학습 방법.

**청구항 20**

제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하고,

상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하고,

상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교차 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는 스텝을 포함하는 처리를 컴퓨터에 실행시키는, 프로그램.

**발명의 설명**

**기술 분야**

[0001] 본 기술은, 신호 처리 장치 및 방법, 학습 장치 및 방법, 그리고 프로그램에 관한 것으로, 특히, 저비용의 장치에서도 고품질의 오디오 재생을 행할 수 있도록 한 신호 처리 장치 및 방법, 학습 장치 및 방법, 그리고 프로그램에 관한 것이다.

**배경 기술**

[0002] 종래, 영화나 게임 등에서 오브젝트 오디오 기술이 사용되고, 오브젝트 오디오를 취급할 수 있는 부호화 방식도 개발되고 있다. 구체적으로는, 예를 들어 국제 표준 규격인 MPEG(Moving Picture Experts Group)-H Part 3: 3D audio 규격 등이 알려져 있다(예를 들어, 비특허문헌 1 참조).

[0003] 이러한 부호화 방식에서는, 종래 2채널 스테레오 방식이나 5.1채널 등의 멀티 채널 스테레오 방식과 함께, 이동하는 음원 등을 독립된 오디오 오브젝트(이하, 단순히 오브젝트라고도 칭함)로서 취급하고, 오디오 오브젝트의 신호 데이터와 함께 오브젝트의 위치 정보를 메타데이터로서 부호화하는 것이 가능하다.

[0004] 이에 의해, 스피커의 수나 배치가 다른 다양한 시청 환경에서 재생을 행할 수 있다. 또한, 종래의 부호화 방식에서는 곤란했던 특정한 음원의 소리의 음량 조정이나, 특정한 음원의 소리에 대한 이펙트의 추가 등, 특정한 음원의 소리를 재생 시에 가공하는 것을 용이하게 할 수 있다.

[0005] 이러한 부호화 방식에서는, 복호측에 있어서 비트 스트림에 대한 디코드가 행해지고, 오브젝트의 오디오 신호인 오브젝트 신호와, 공간 내에 있어서의 오브젝트의 위치를 나타내는 오브젝트 위치 정보를 포함하는 메타데이터가 얻어진다.

[0006] 그리고, 오브젝트 위치 정보에 기초하여, 공간 내에 가상적으로 배치된 복수의 각 가상 스피커에 오브젝트 신호를 렌더링하는 렌더링 처리가 행해진다. 예를 들어, 비특허문헌 1의 규격에서는, 렌더링 처리에 3차원 VBAP(Vector Based Amplitude Panning)(이하, 단순히 VBAP라고 칭함)라고 불리는 방식이 사용된다.

[0007] 또한, 렌더링 처리에 의해, 각 가상 스피커에 대응하는 가상 스피커 신호가 얻어지면, 그것들의 가상 스피커 신호에 기초하여 HRTF(Head Related Transfer Function) 처리가 행해진다. 이 HRTF 처리에서는, 마치 가상 스피커로부터 소리가 재생되고 있는 것처럼 실제의 헤드폰이나 스피커로부터 소리를 출력시키기 위한 출력 오디오 신호가 생성된다.

[0008] 이러한 오브젝트 오디오를 실제로 재생하는 경우, 공간 위에 실제의 스피커를 다수 배치할 수 있을 때에는, 가상 스피커 신호에 기초하는 재생이 행해진다. 또한, 다수의 스피커를 배치할 수 없어, 헤드폰이나 사운드바 등의 소수의 스피커에서 오브젝트 오디오를 재생할 때에는, 상술한 출력 오디오 신호에 기초하는 재생이 행해진다.

- [0009] 한편, 근년, 스토리지 가격의 하락이나 네트워크의 광대역화에 의해, 샘플링 주파수가 96kHz 이상인, 소위 하이 레조 음원, 즉 하이 레졸루션 음원을 즐길 수 있도록 되어 오고 있다.
- [0010] 비특허문헌 1에 기재된 부호화 방식에서는, 하이 레조 음원을 효율적으로 부호화하기 위한 기술로서, SBR(Spectral Band Replication) 등의 기술을 사용할 수 있다.
- [0011] 예를 들어 SBR에 있어서는, 부호화측에서는, 스펙트럼의 고역 성분은 부호화되지 않고, 고역 서브 밴드 신호의 평균 진폭 정보가 고역 서브 밴드의 개수분만큼 부호화되어 전송된다.
- [0012] 그리고, 복호측에 있어서는, 저역 서브 밴드 신호와, 고역의 평균 진폭 정보에 기초하여, 저역 성분과 고역 성분이 포함되는 최종적인 출력 신호가 생성된다. 이에 의해, 더 고품질의 오디오 재생을 실현할 수 있다.
- [0013] 이 방법에서는, 인간은 고역 신호 성분의 위상 변화에는 둔감하고, 그 주파수 포락의 개형이 원래의 신호에 가까운 경우, 그 차를 지각할 수 없다는 청각 특성이 이용되고 있고, 이러한 방법은, 일반적으로 대역 확장 기술로서 널리 알려져 있다.

**선행기술문헌**

**비특허문헌**

- [0014] (비특허문헌 0001) INTERNATIONAL STANDARD ISO/IEC 23008-3 Second edition 2019-02 Information technology-High efficiency coding and media delivery in heterogeneous environments-Part 3: 3D audio

**발명의 내용**

**해결하려는 과제**

- [0015] 그런데, 상술한 오브젝트 오디오에 대하여, 렌더링 처리나 HRTF 처리와 조합하여 대역 확장을 행하는 경우, 각 오브젝트의 오브젝트 신호에 대하여 대역 확장 처리가 행해지고 나서, 렌더링 처리나 HRTF 처리가 행해진다.
- [0016] 이 경우, 대역 확장 처리는 오브젝트의 수만큼 독립적으로 행해지기 때문에, 처리 부하, 즉 연산량이 많아져 버린다. 또한, 대역 확장 처리 후에는, 대역 확장에 의해 얻어진, 샘플링 주파수가 더 높은 신호를 대상으로 하여 렌더링 처리나 HRTF 처리가 행해지기 때문에, 처리 부하가 더 증대해 버린다.
- [0017] 그렇게 하면, 저비용의 프로세서나 배터리의 장치, 즉 연산 처리 능력이 낮은 장치나, 배터리 용량이 적은 장치 등, 저비용의 장치에서는 대역 확장을 행할 수 없어, 결과적으로 고품질의 오디오 재생을 행할 수 없게 되어 버린다.
- [0018] 본 기술은, 이러한 상황을 감안하여 이루어진 것이고, 저비용의 장치에서도 고품질의 오디오 재생을 행할 수 있도록 하는 것이다.

**과제의 해결 수단**

- [0019] 본 기술의 제1 측면의 신호 처리 장치는, 입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하는 디코드 처리부와, 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 대역 확장부를 구비한다.
- [0020] 본 기술의 제1 측면의 신호 처리 방법 또는 프로그램은, 입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하고, 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 스텝을 포함한다.
- [0021] 본 기술의 제1 측면에 있어서는, 입력 비트 스트림이 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화되고, 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리가 행해져, 출력 오디오 신호가 생성된다.

[0022] 본 기술의 제2 측면의 학습 장치는, 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하는 제1 고역 정보 계산부와, 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하는 제2 고역 정보 계산부와, 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는 고역 정보 학습부를 구비한다.

[0023] 본 기술의 제2 측면의 학습 방법 또는 프로그램은, 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하고, 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하고, 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는 스텝을 포함한다.

[0024] 본 기술의 제2 측면에 있어서는, 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보가 생성되고, 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보가 생성되고, 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습이 행해져, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터가 생성된다.

**도면의 간단한 설명**

- [0025] 도 1은 출력 오디오 신호의 생성에 대하여 설명하는 도면이다.
- 도 2는 VBAP에 대하여 설명하는 도면이다.
- 도 3은 HRTF 처리에 대하여 설명하는 도면이다.
- 도 4는 대역 확장 처리에 대하여 설명하는 도면이다.
- 도 5는 대역 확장 처리에 대하여 설명하는 도면이다.
- 도 6은 신호 처리 장치의 구성예를 나타내는 도면이다.
- 도 7은 본 기술을 적용한 신호 처리 장치의 구성예를 나타내는 도면이다.
- 도 8은 개인용 고역 정보 생성부의 구성예를 나타내는 도면이다.
- 도 9는 입력 비트 스트림의 선택예를 나타내는 도면이다.
- 도 10은 신호 생성 처리를 설명하는 흐름도이다.
- 도 11은 학습 장치의 구성예를 나타내는 도면이다.
- 도 12는 학습 처리를 설명하는 흐름도이다.
- 도 13은 인코더의 구성예를 나타내는 도면이다.
- 도 14는 부호화 처리를 설명하는 흐름도이다.
- 도 15는 컴퓨터의 구성예를 나타내는 도면이다.

**발명을 실시하기 위한 구체적인 내용**

[0026] 이하, 도면을 참조하여, 본 기술을 적용한 실시 형태에 대하여 설명한다.

[0027] <제1 실시 형태>

[0028] <본 기술에 대하여>

[0029] 본 기술은, 미리 HRTF 출력 신호를 대상으로 한 대역 확장 처리를 위한 범용의 고역 정보를 비트 스트림 중에

다중화하여 전송함과 함께, 복호측에 있어서, 개인용의 HRTF 계수와, 범용의 HRTF 계수 및 고역 정보에 기초하여, 개인용의 HRTF 계수에 대응하는 고역 정보를 생성하도록 했다.

- [0030] 이에 의해, 처리 부하가 높은 디코드 처리나 렌더링 처리, 버추얼라이즈 처리를 낮은 샘플링 주파수에서 행하고, 그 후, 개인용의 HRTF 계수에 대응하는 고역 정보에 기초하여 대역 확장 처리를 행하도록 할 수 있어, 전체적으로 연산량을 저감시킬 수 있다. 그 결과, 저비용의 장치에서도, 더 높은 샘플링 주파수의 출력 오디오 신호에 기초하는, 고품질의 오디오 재생을 행할 수 있다.
- [0031] 특히, 본 기술에서는, 복호측에서 개인용의 HRTF 계수에 대응하는 고역 정보를 생성함으로써, 부호화측에서 유저 개인마다 고역 정보를 준비할 필요가 없어진다. 또한, 복호측에서는 개인용의 HRTF 계수에 대응하는 고역 정보를 생성함으로써, 범용의 고역 정보를 사용하는 경우보다도, 더 고품질의 오디오 재생을 행할 수 있다.
- [0032] 그러면, 이하, 본 기술에 대하여, 더 상세하게 설명한다.
- [0033] 먼저, MPEG-H Part 3: 3D audio 규격의 부호화 방식에서의 부호화에 의해 얻어진 비트 스트림에 대하여 복호(디코드)를 행하여, 오브젝트 오디오의 출력 오디오 신호를 생성할 때 행해지는 일반적인 처리에 대하여 설명한다.
- [0034] 예를 들어 도 1에 나타난 바와 같이, 부호화(인코드)에 의해 얻어진 입력 비트 스트림이 디코드 처리부(11)에 입력되면, 그 입력 비트 스트림에 대하여 비다중화 및 디코드 처리가 행해진다.
- [0035] 디코드 처리에 의해, 콘텐츠를 구성하는 오브젝트(오디오 오브젝트)의 소리를 재생하기 위한 오디오 신호인 오브젝트 신호와, 그 오브젝트의 공간 내의 위치를 나타내는 오브젝트 위치 정보를 포함하는 메타데이터가 얻어진다.
- [0036] 계속해서, 렌더링 처리부(12)에서는, 메타데이터에 포함되는 오브젝트 위치 정보에 기초하여, 공간 내에 가상적으로 배치된 가상 스피커에 오브젝트 신호를 렌더링하는 렌더링 처리가 행해지고, 각 가상 스피커로부터 출력되는 소리를 재생하기 위한 가상 스피커 신호가 생성된다.
- [0037] 또한, 버추얼라이즈 처리부(13)에서는, 각 가상 스피커의 가상 스피커 신호에 기초하여 버추얼라이즈 처리가 행해지고, 유저가 장착하는 헤드폰이나 실공간에 배치된 스피커 등의 재생 장치로부터 소리를 출력시키기 위한 출력 오디오 신호가 생성된다.
- [0038] 버추얼라이즈 처리란, 실제의 재생 환경에서의 채널 구성과는 다른 채널 구성으로 재생이 행해지고 있는 오디오 재생을 실현하기 위한 오디오 신호를 생성하는 처리이다.
- [0039] 예를 들어, 이 예에서는, 실제로는 헤드폰 등의 재생 장치로부터 소리가 출력되고 있음에도 불구하고, 마치 각 가상 스피커로부터 소리가 출력되고 있는 것과 같은 오디오 재생을 실현하기 위한 출력 오디오 신호를 생성하는 처리가 버추얼라이즈 처리이다.
- [0040] 버추얼라이즈 처리는, 어떤 방법에 의해 실현되어도 되지만, 이하에는 버추얼라이즈 처리로서 HRTF 처리가 행해지는 것으로서 설명을 계속한다.
- [0041] 버추얼라이즈 처리에서 얻어진 출력 오디오 신호에 기초하여, 실제의 헤드폰이나 스피커로부터 소리를 출력하면, 마치 가상 스피커로부터 소리가 재생되고 있는 것과 같은 오디오 재생을 실현할 수 있다. 또한, 이하에는, 실공간에 실제로 배치되는 스피커를 특히 실제 스피커라고도 칭하는 것으로 한다.
- [0042] 이러한 오브젝트 오디오를 재생하는 경우, 공간 내에 다수의 실제 스피커를 배치할 수 있을 때는, 렌더링 처리의 출력을 그대로 실제 스피커에서 재생할 수 있다.
- [0043] 이에 비해, 공간 내에 다수의 실제 스피커를 배치할 수 없을 때는, HRTF 처리를 행하여 헤드폰이나, 사운드바 등의 소수의 실제 스피커에 의해 재생을 행하게 된다. 일반적으로는, 헤드폰이나 소수의 실제 스피커에 의해 재생을 행하는 경우가 많다.
- [0044] 여기서, 일반적인 렌더링 처리와 HRTF 처리에 대하여, 더 설명을 행한다.
- [0045] 예를 들어 렌더링 시에는, 상술한 VBAP 등의 소정의 방식의 렌더링 처리가 행해진다. VBAP은 일반적으로 패닝이라고 불리는 렌더링 방법의 하나로, 유저 위치를 원점으로 하는 구 표면 상에 존재하는 가상 스피커 중, 동일하게 구 표면 상에 존재하는 오브젝트에 가장 가까운 3개의 가상 스피커에 대하여 계인을 분배함으로써 렌더링을 행하는 것이다.
- [0046] 예를 들어 도 2에 나타난 바와 같이, 3차원 공간에 수평자인 유저 U11이 있고, 그 유저 U11의 전방에 3개의 가

상 스피커 SP1 내지 가상 스피커 SP3이 배치되어 있는 것으로 한다.

[0047] 여기서, 유저 U11의 헤드부의 위치를 원점 0라고 하고, 그 원점 0를 중심으로 하는 구의 표면 상에 가상 스피커 SP1 내지 가상 스피커 SP3이 위치하고 있는 것으로 한다.

[0048] 지금, 구 표면 상에 있어서의 가상 스피커 SP1 내지 가상 스피커 SP3에 둘러싸이는 영역 TR11 내에 오브젝트가 존재하고 있고, 그 오브젝트의 위치 VSP1에 음상을 정위시키는 것을 생각하는 것으로 한다.

[0049] 그러한 경우, VBAP에서는 오브젝트에 대하여, 위치 VSP1의 주위에 있는 가상 스피커 SP1 내지 가상 스피커 SP3에 대하여 게인이 분배되게 된다.

[0050] 구체적으로는, 원점 0를 기준(원점)으로 하는 3차원 좌표계에 있어서, 원점 0를 시점이라고 하고, 위치 VSP1을 종점으로 하는 3차원의 벡터 P에 의해 위치 VSP1을 나타내는 것으로 한다.

[0051] 또한, 원점 0를 시점이라고 하고, 각 가상 스피커 SP1 내지 가상 스피커 SP3의 위치를 종점으로 하는 3차원의 벡터를 벡터 L<sub>1</sub> 내지 벡터 L<sub>3</sub>으로 하면, 벡터 P는 다음 식 (1)에 나타낸 바와 같이 벡터 L<sub>1</sub> 내지 벡터 L<sub>3</sub>의 선형 합에 의해 나타낼 수 있다.

$$P = g_1 L_1 + g_2 L_2 + g_3 L_3 \quad \dots (1)$$

[0053] 여기서, 식 (1)에 있어서 벡터 L<sub>1</sub> 내지 벡터 L<sub>3</sub>에 승산되어 있는 계수 g<sub>1</sub> 내지 계수 g<sub>3</sub>을 산출하고, 이들 계수 g<sub>1</sub> 내지 계수 g<sub>3</sub>을, 가상 스피커 SP1 내지 가상 스피커 SP3의 각각으로부터 출력하는 소리의 게인으로 하면, 위치 VSP1에 음상을 정위시킬 수 있다.

[0054] 예를 들어, 계수 g<sub>1</sub> 내지 계수 g<sub>3</sub>을 요소로 하는 벡터를 g<sub>123</sub>=[g<sub>1</sub>, g<sub>2</sub>, g<sub>3</sub>]이라고 하고, 벡터 L<sub>1</sub> 내지 벡터 L<sub>3</sub>을 요소로 하는 벡터를 L<sub>123</sub>=[L<sub>1</sub>, L<sub>2</sub>, L<sub>3</sub>]으로 하면, 상술한 식 (1)을 변형하여 다음 식 (2)를 얻을 수 있다.

$$g_{123} = P^T L_{123}^{-1} \quad \dots (2)$$

[0056] 이러한 식 (2)를 계산하여 구한 계수 g<sub>1</sub> 내지 계수 g<sub>3</sub>을 게인으로 사용하여, 오브젝트 신호에 기초하는 소리를 각 가상 스피커 SP1 내지 가상 스피커 SP3으로부터 출력하면, 위치 VSP1에 음상을 정위시킬 수 있다.

[0057] 또한, 각 가상 스피커 SP1 내지 가상 스피커 SP3의 배치 위치는 고정되어 있고, 그것들의 가상 스피커의 위치를 나타내는 정보는 기지이기 때문에, 역행렬인 L<sub>123</sub><sup>-1</sup>은 사전에 구해 둘 수 있다.

[0058] 도 2에 나타낸 구 표면 상에 있어서의, 3개의 가상 스피커에 의해 둘러싸이는 삼각형의 영역 TR11은 메쉬라고 불리고 있다. 공간 내에 배치된 다수의 가상 스피커를 조합하여 복수의 메쉬를 구성함으로써, 오브젝트의 소리를 공간 내의 임의의 위치에 정위시키는 것이 가능하다.

[0059] 이와 같이, 각 오브젝트에 대하여 가상 스피커의 게인이 구해지면, 다음 식 (3)의 연산을 행함으로써, 각 가상 스피커의 가상 스피커 신호를 얻을 수 있다.

$$\begin{bmatrix} SP(0, t) \\ SP(1, t) \\ \vdots \\ SP(M-1, t) \end{bmatrix} = \begin{bmatrix} G(0, 0) & G(0, 1) & \dots & G(0, N-1) \\ G(1, 0) & G(1, 1) & \dots & G(1, N-1) \\ \vdots & \vdots & & \vdots \\ G(M-1, 0) & G(M-1, 1) & \dots & G(M-1, N-1) \end{bmatrix} \begin{bmatrix} S(0, t) \\ S(1, t) \\ \vdots \\ S(N-1, t) \end{bmatrix} \quad \dots (3)$$

[0061] 또한, 식 (3)에 있어서 SP(m, t)는, M개의 가상 스피커 중 m번째(단, m=0, 1, ..., M-1)의 가상 스피커의 시각 t에 있어서의 가상 스피커 신호를 나타내고 있다. 또한, 식 (3)에 있어서 S(n, t)는 N개의 오브젝트 중 n번째(단, n=0, 1, ..., N-1)의 오브젝트의 시각 t에 있어서의 오브젝트 신호를 나타내고 있다.

[0062] 또한 식 (3)에 있어서 G(m, n)는, m번째의 가상 스피커에 대한 가상 스피커 신호 SP(m, t)를 얻기 위한, n번째

의 오브젝트의 오브젝트 신호  $S(n, t)$ 에 승산되는 계인을 나타내고 있다. 즉, 계인  $G(m, n)$ 는, 상술한 식 (2)에 의해 구해진,  $n$ 번째의 오브젝트에 대한  $m$ 번째의 가상 스피커에 분배된 계인을 나타내고 있다.

- [0063] 렌더링 처리에서는, 이 식 (3)의 계산이 가장 계산 비용이 드는 처리가 된다. 즉, 식 (3)의 연산이 가장 연산량이 많은 처리가 된다.
- [0064] 이어서, 식 (3)의 연산에 의해 얻어진 가상 스피커 신호에 기초하는 소리를 헤드폰 또는 소수의 실제 스피커에서 재생하는 경우에 행해지는 HRTF 처리의 예에 대하여 도 3을 참조하여 설명한다. 또한, 도 3에서는 설명을 간단하게 하기 위해, 2차원의 수평면 상에 가상 스피커가 배치된 예로 되어 있다.
- [0065] 도 3에서는, 공간 내에 5개의 가상 스피커 SP11-1 내지 가상 스피커 SP11-5가 원 형상으로 배열되어 배치되어 있다. 이하, 가상 스피커 SP11-1 내지 가상 스피커 SP11-5를 특별히 구별할 필요가 없는 경우, 단순히 가상 스피커 SP11이라고도 칭하는 것으로 한다.
- [0066] 또한, 도 3에서는 5개의 가상 스피커 SP11에 둘러싸이는 위치, 즉 가상 스피커 SP11이 배치된 원의 중심 위치에 수청자인 유저 U21이 위치하고 있다. 따라서, HRTF 처리에서는, 마치 유저 U21이 각 가상 스피커 SP11로부터 출력되는 소리를 듣고 있는 것과 같은 오디오 재생을 실현하기 위한 출력 오디오 신호가 생성된다.
- [0067] 특히, 이 예에서는 유저 U21이 있는 위치를 청취 위치로 하고, 5개의 각 가상 스피커 SP11로의 렌더링에 의해 얻어진 가상 스피커 신호에 기초하는 소리를 헤드폰에 의해 재생하는 것으로 한다.
- [0068] 그러한 경우, 예를 들어 가상 스피커 신호에 기초하여 가상 스피커 SP11-1로부터 출력(방사)된 소리는 화살표 Q11에 나타내는 경로를 통해, 유저 U21의 좌귀의 고막에 도달한다. 그 때문에, 가상 스피커 SP11-1로부터 출력된 소리의 특성은, 가상 스피커 SP11-1로부터 유저 U21의 좌귀까지의 공간 전달 특성, 유저 U21의 얼굴이나 귀의 형상이나 반사 흡수 특성 등에 의해 변화될 것이다.
- [0069] 그래서, 가상 스피커 SP11-1의 가상 스피커 신호에 대하여, 가상 스피커 SP11-1로부터 유저 U21의 좌귀까지의 공간 전달 특성 및 유저 U21의 얼굴이나 귀의 형상, 반사 흡수 특성 등이 가미된 전달 함수  $H_{L\_SP11}$ 을 컨볼루션하면, 유저 U21의 좌귀에서 들릴 가상 스피커 SP11-1로부터의 소리를 재생하는 출력 오디오 신호를 얻을 수 있다.
- [0070] 마찬가지로, 예를 들어 가상 스피커 신호에 기초하여 가상 스피커 SP11-1로부터 출력된 소리는 화살표 Q12에 나타내는 경로를 통해, 유저 U21의 우귀의 고막에 도달한다. 따라서, 가상 스피커 SP11-1의 가상 스피커 신호에 대하여, 가상 스피커 SP11-1로부터 유저 U21의 우귀까지의 공간 전달 특성 및 유저 U21의 얼굴이나 귀의 형상, 반사 흡수 특성 등이 가미된 전달 함수  $H_{R\_SP11}$ 을 컨볼루션하면, 유저 U21의 우귀에서 들릴 가상 스피커 SP11-1로부터의 소리를 재생하는 출력 오디오 신호를 얻을 수 있다.
- [0071] 이러한 점에서, 최종적으로 5개의 가상 스피커 SP11의 가상 스피커 신호에 기초하는 소리를 헤드폰에서 재생할 때는, 좌채널에 대해서는, 각 가상 스피커 신호에 대하여, 각 가상 스피커의 좌귀용의 전달 함수를 컨볼루션하고, 그 결과 얻어진 각 신호를 더하여 좌채널의 출력 오디오 신호로 하면 된다.
- [0072] 마찬가지로, 우채널에 대해서는, 각 가상 스피커 신호에 대하여, 각 가상 스피커의 우귀용의 전달 함수를 컨볼루션하고, 그 결과 얻어진 각 신호를 더하여 우채널의 출력 오디오 신호로 하면 된다.
- [0073] 또한, 재생에 사용하는 재생 장치가 헤드폰이 아니라 실제 스피커인 경우에도, 헤드폰에 있어서의 경우와 마찬가지로의 HRTF 처리가 행해진다. 그러나, 이 경우에는 스피커로부터의 소리는 공간 전반에 의해 유저의 좌우의 양쪽 귀에 도달하기 때문에, 크로스토크가 고려된 처리가 행해지게 된다. 이러한 처리는 트랜스 오럴 처리라고도 불리고 있다.
- [0074] 일반적으로는 주파수 표현된 좌귀용, 즉 좌채널의 출력 오디오 신호를  $L(\omega)$ 이라고 하고, 주파수 표현된 우귀용, 즉 우채널의 출력 오디오 신호를  $R(\omega)$ 이라고 하면, 이들  $L(\omega)$  및  $R(\omega)$ 은 다음 식 (4)를 계산함으로써 얻을 수 있다.

$$\begin{bmatrix} L(\omega) \\ R(\omega) \end{bmatrix} = \begin{bmatrix} H_{L}(0, \omega) & H_{L}(1, \omega) & \cdots & H_{L}(M-1, \omega) \\ H_{R}(0, \omega) & H_{R}(1, \omega) & \cdots & H_{R}(M-1, \omega) \end{bmatrix} \begin{bmatrix} SP(0, \omega) \\ SP(1, \omega) \\ \vdots \\ SP(M-1, \omega) \end{bmatrix}$$

... (4)

- [0075]
- [0076] 또한, 식 (4)에 있어서  $\omega$ 는 주파수를 나타내고 있고,  $SP(m, \omega)$ 는 M개의 가상 스피커 중 m번째(단,  $m=0, 1, \dots, M-1$ )의 가상 스피커의 주파수  $\omega$ 의 가상 스피커 신호를 나타내고 있다. 가상 스피커 신호  $SP(m, \omega)$ 는, 상술한 가상 스피커 신호  $SP(m, t)$ 를 시간 주파수 변환함으로써 얻을 수 있다.
- [0077] 또한, 식 (4)에 있어서  $H_L(m, \omega)$ 은, 좌채널의 출력 오디오 신호  $L(\omega)$ 을 얻기 위한, m번째의 가상 스피커에 대한 가상 스피커 신호  $SP(m, \omega)$ 에 승산되는 좌귀용의 전달 함수를 나타내고 있다. 마찬가지로  $H_R(m, \omega)$ 은 우귀용의 전달 함수를 나타내고 있다.
- [0078] 이들 HRTF의 전달 함수  $H_L(m, \omega)$ 이나 전달 함수  $H_R(m, \omega)$ 을 시간 영역의 임펄스 응답으로서 표현하는 경우, 적어도 1초 정도의 길이가 필요해진다. 그 때문에, 예를 들어 가상 스피커 신호의 샘플링 주파수가 48kHz인 경우에는, 48000 탭의 컨벌루션을 행해야만 해, 전달 함수의 컨벌루션에 FFT(Fast Fourier Transform)를 사용한 고속 연산 방법을 사용해도 또한 많은 연산량이 필요해진다.
- [0079] 이상과 같이 디코드 처리, 렌더링 처리 및 HRTF 처리를 행하여 출력 오디오 신호를 생성하고, 헤드폰이나 소수 개의 실제 스피커를 사용하여 오브젝트 오디오를 재생하는 경우, 많은 연산량이 필요해진다. 또한, 이 연산량은 오브젝트의 수가 증가하면, 그만큼 더 많아진다.
- [0080] 이어서, 대역 확장 처리에 대하여 설명한다.
- [0081] 일반적인 대역 확장 처리, 즉 SBR에서는, 부호화측에 있어서, 오디오 신호의 스펙트럼의 고역 성분은 부호화되지 않고, 고역의 주파수 대역인 고역 서브 밴드의 고역 서브 밴드 신호의 평균 진폭 정보가 고역 서브 밴드의 개수분 부호화되어, 복호측으로 전송된다.
- [0082] 또한, 복호측에서는, 디코드 처리(복호)에 의해 얻어진 오디오 신호인 저역 서브 밴드 신호가, 그 평균 진폭으로 정규화된 후, 정규화된 신호가 고역 서브 밴드로 카피(복제)된다. 그리고, 그 결과 얻어진 신호에 각 고역 서브 밴드의 평균 진폭 정보가 승산되어 고역 서브 밴드 신호로 되고, 저역 서브 밴드 신호와 고역 서브 밴드 신호가 서브 밴드 합성되어, 최종적인 출력 오디오 신호로 된다.
- [0083] 이러한 대역 확장 처리에 의해, 예를 들어 샘플링 주파수가 96kHz 이상인 하이 레조 음원의 오디오 재생을 행할 수 있다.
- [0084] 그러나, 예를 들어 일반적인 스테레오의 오디오와는 달리, 오브젝트 오디오에 있어서 샘플링 주파수가 96kHz인 신호를 처리하는 경우, SBR 등의 대역 확장 처리가 행해지는지 여부에 구애되지 않고, 복호에 의해 얻어진 96kHz의 오브젝트 신호에 대하여, 렌더링 처리나 HRTF 처리가 행해지게 된다. 그 때문에, 오브젝트수나 가상 스피커 수가 많은 경우, 그것들의 처리의 계산 비용은 막대한 것으로 되어, 고성능의 프로세서와 높은 소비 전력이 필요해진다.
- [0085] 여기서, 도 4를 참조하여, 오브젝트 오디오에 있어서 대역 확장에 의해 96kHz의 출력 오디오 신호를 얻는 경우에 행해지는 처리에 대하여 설명한다. 또한, 도 4에 있어서 도 1에 있어서의 경우와 대응하는 부분에는 동일한 부호를 부여하고 있고, 그 설명은 생략한다.
- [0086] 입력 비트 스트림이 공급되면, 디코드 처리부(11)에서 비다중화 및 디코드 처리가 행해지고, 그 결과 얻어진 오브젝트 신호와, 오브젝트의 오브젝트 위치 정보 및 고역 정보가 출력된다.
- [0087] 예를 들어 고역 정보는, 부호화 전의 오브젝트 신호로부터 얻어지는 고역 서브 밴드 신호의 평균 진폭 정보이다.
- [0088] 바꾸어 말하면 고역 정보는, 디코드 처리에서 얻어지는 오브젝트 신호에 대응하는, 샘플링 주파수가 더 높은 부호화 전의 오브젝트 신호의 고역측의 각 서브 밴드 성분의 크기를 나타내는, 대역 확장을 위한 대역 확장 정보이다. 또한, 여기서는 SBR을 예로서 설명을 행하고 있기 때문에, 대역 확장 정보로서 고역 서브 밴드 신호의

평균 진폭 정보가 사용되고 있지만, 대역 확장 처리를 위한 대역 확장 정보는, 부호화 전의 오브젝트 신호의 고역측의 각 서브 밴드의 진폭의 대푯값이나, 주파수 포락의 형상을 나타내는 정보 등, 어떤 것이어도 된다.

- [0089] 또한, 여기서는, 디코드 처리에 의해 얻어지는 오브젝트 신호는, 예를 들어 샘플링 주파수가 48kHz인 것으로 하고, 이하에는, 그러한 오브젝트 신호를 저FS 오브젝트 신호라고도 칭하는 것으로 한다.
- [0090] 디코드 처리 후, 대역 확장부(41)에서는, 고역 정보와 저FS 오브젝트 신호에 기초하여 대역 확장 처리가 행해져, 샘플링 주파수가 더 높은 오브젝트 신호가 얻어진다. 이 예에서는, 대역 확장 처리에 의해, 예를 들어 샘플링 주파수가 96kHz인 오브젝트 신호가 얻어지는 것으로 하고, 이하에는, 그러한 오브젝트 신호를 고FS 오브젝트 신호라고도 칭하는 것으로 한다.
- [0091] 또한, 렌더링 처리부(12)에서는, 디코드 처리에 의해 얻어진 오브젝트 위치 정보와, 대역 확장 처리에 의해 얻어진 고FS 오브젝트 신호에 기초하여 렌더링 처리가 행해진다. 특히, 이 예에서는 렌더링 처리에 의해, 샘플링 주파수가 96kHz인 가상 스피커 신호가 얻어지고, 이하에는, 그러한 가상 스피커 신호를 고FS 가상 스피커 신호라고도 칭한다.
- [0092] 또한, 그 후, 버추얼라이즈 처리부(13)에 있어서, 고FS 가상 스피커 신호에 기초하여 HRTF 처리 등의 버추얼라이즈 처리가 행해져, 샘플링 주파수가 96kHz인 출력 오디오 신호가 얻어진다.
- [0093] 여기서, 도 5를 참조하여, 일반적인 대역 확장 처리에 대하여 설명한다.
- [0094] 도 5는, 소정의 오브젝트 신호의 주파수 진폭 특성을 나타내고 있다. 또한, 도 5에 있어서 종축은 진폭(파워)을 나타내고 있고, 횡축은 주파수를 나타내고 있다.
- [0095] 예를 들어 꺾은선 L11은, 대역 확장부(41)에 공급되는 저FS 오브젝트 신호의 주파수 진폭 특성을 나타내고 있다. 이 저FS 오브젝트 신호는, 샘플링 주파수가 48kHz이고, 저FS 오브젝트 신호에는 24kHz 이상의 주파수 대역의 신호 성분은 포함되어 있지 않다.
- [0096] 여기서는, 예를 들어 24kHz까지의 주파수 대역이, 저역 서브 밴드 sb-8 내지 저역 서브 밴드 sb-1을 포함하는 복수의 저역 서브 밴드로 분할되어 있고, 그것들의 각 저역 서브 밴드의 신호 성분이 저역 서브 밴드 신호이다. 마찬가지로, 24kHz부터 48kHz까지의 주파수 대역이, 고역 서브 밴드 sb 내지 고역 서브 밴드 sb+13으로 분할되어 있고, 그것들의 각 고역 서브 밴드의 신호 성분이 고역 서브 밴드 신호이다.
- [0097] 또한, 대역 확장부(41)에는, 각 고역 서브 밴드 sb 내지 고역 서브 밴드 sb+13에 대하여, 그것들의 고역 서브 밴드의 평균 진폭 정보를 나타내는 고역 정보가 공급된다.
- [0098] 예를 들어 도 5에서는, 직선 L12는, 고역 서브 밴드 sb의 고역 정보로서 공급되는 평균 진폭 정보를 나타내고 있고, 직선 L13은, 고역 서브 밴드 sb+1의 고역 정보로서 공급되는 평균 진폭 정보를 나타내고 있다.
- [0099] 대역 확장부(41)에서는, 저역 서브 밴드 신호가, 그 저역 서브 밴드 신호의 평균 진폭값으로 정규화되고, 정규화에 의해 얻어진 신호가 고역측으로 카피(매핑)된다. 여기서, 카피원이 되는 저역 서브 밴드와, 그 저역 서브 밴드의 카피처가 되는 고역 서브 밴드는 확장 주파수 대역 등에 의해 미리 정해져 있다.
- [0100] 예를 들어 저역 서브 밴드 sb-8의 저역 서브 밴드 신호가 정규화되고, 정규화에 의해 얻어진 신호가, 고역 서브 밴드 sb로 카피된다.
- [0101] 더 구체적으로는, 저역 서브 밴드 sb-8의 저역 서브 밴드 신호의 정규화 후의 신호에 대하여 변조 처리가 행해져, 고역 서브 밴드 sb의 주파수 성분의 신호로 변환된다.
- [0102] 마찬가지로, 예를 들어 저역 서브 밴드 sb-7의 저역 서브 밴드 신호는, 정규화 후, 고역 서브 밴드 sb+1로 카피된다.
- [0103] 이와 같이 하여 정규화된 저역 서브 밴드 신호가 고역 서브 밴드로 카피(매핑)되면, 각 고역 서브 밴드의 카피된 신호에 대하여, 그것들의 각 고역 서브 밴드의 고역 정보에 의해 나타나는 평균 진폭 정보가 승산되어, 고역 서브 밴드 신호가 생성된다.
- [0104] 예를 들어 고역 서브 밴드 sb에서는, 저역 서브 밴드 sb-8의 저역 서브 밴드 신호를 정규화하여 고역 서브 밴드 sb로 카피함으로써 얻어진 신호에 대하여, 직선 L12에 의해 나타나는 평균 진폭 정보가 승산되어, 고역 서브 밴드 sb의 고역 서브 밴드 신호로 된다.
- [0105] 각 고역 서브 밴드에 대하여, 고역 서브 밴드 신호가 얻어지면, 그 후, 각 저역 서브 밴드 신호와, 각 고역 서

브 밴드 신호가 96kHz 샘플링의 대역 합성 필터에 입력되어 필터링(합성)되고, 그 결과 얻어진 고FS 오브젝트 신호가 출력된다. 즉, 샘플링 주파수가 96kHz로 업 샘플링(대역 확장)된, 고FS 오브젝트 신호가 얻어진다.

- [0106] 도 4에 나타낸 예에서는, 대역 확장부(41)에 있어서, 이상과 같은 고FS 오브젝트 신호를 생성하는 대역 확장 처리가, 입력 비트 스트림에 포함되는 저FS 오브젝트 신호마다, 즉 오브젝트마다 독립되어 행해진다.
- [0107] 따라서, 예를 들어 오브젝트수가 32개인 경우, 렌더링 처리부(12)에서는, 32개의 각 오브젝트에 대하여, 96kHz의 고FS 오브젝트 신호의 렌더링 처리를 행해야만 한다.
- [0108] 마찬가지로, 그 후단의 버추얼라이즈 처리부(13)에 있어서도, 가상 스피커 수분만큼, 96kHz의 고FS 가상 스피커 신호의 HRTF 처리(버추얼라이즈 처리)를 행해야만 한다.
- [0109] 그 결과, 장치 전체에 있어서의 처리 부하는 막대한 것으로 되어 버린다. 이것은, 대역 확장 처리를 행하지 않고, 디코드 처리에 의해 얻어지는 오디오 신호의 샘플링 주파수가 96kHz인 경우라도 마찬가지이다.
- [0110] 그래서, 하이 레조, 즉 높은 샘플링 주파수의 버추얼라이즈 처리 후의 신호의 고역 정보를 부호화 시에 미리 계산하고, 입력 비트 스트림에 다중화하여 전송하는 것이 생각된다.
- [0111] 이렇게 함으로써, 예를 들어 처리 부하가 높은 디코드 처리, 렌더링 처리, HRTF 처리를 낮은 샘플링 주파수에서 행하고, HRTF 처리 후의 최종적인 신호에 대하여, 전송된 고역 정보에 기초하는 대역 확장 처리를 행할 수 있다. 이에 의해, 전체에 있어서의 처리 부하를 저감시켜, 저비용의 프로세서나 배터리에서도 고품질의 오디오 재생을 실현할 수 있다.
- [0112] 그러한 경우, 복호측의 신호 처리 장치를, 예를 들어 도 6에 나타내는 구성으로 할 수 있다. 또한, 도 6에 있어서 도 4에 있어서의 경우와 대응하는 부분에는 동일한 부호를 부여하고 있고, 그 설명은 적절히 생략한다.
- [0113] 도 6에 나타내는 신호 처리 장치(71)는, 예를 들어 스마트폰이나 퍼스널 컴퓨터 등을 포함하고, 디코드 처리부(11), 렌더링 처리부(12), 버추얼라이즈 처리부(13) 및 대역 확장부(41)를 갖고 있다.
- [0114] 도 4에 나타낸 예에서는 디코드 처리, 대역 확장 처리, 렌더링 처리 및 버추얼라이즈 처리의 순으로 각 처리가 행해진다.
- [0115] 이에 비해, 신호 처리 장치(71)에서는, 디코드 처리, 렌더링 처리, 버추얼라이즈 처리 및 대역 확장 처리의 순으로 각 처리(신호 처리)가 행해진다. 즉, 대역 확장 처리가 마지막으로 행해진다.
- [0116] 따라서 신호 처리 장치(71)에서는, 먼저 디코드 처리부(11)에 있어서 입력 비트 스트림의 비다중화 및 디코드 처리가 행해진다.
- [0117] 디코드 처리부(11)는, 비다중화 및 디코드 처리(복호 처리)에 의해 얻어진 고역 정보를 대역 확장부(41)에 공급함과 함께, 오브젝트 위치 정보 및 오브젝트 신호를 렌더링 처리부(12)에 공급한다.
- [0118] 여기서, 입력 비트 스트림에는, 버추얼라이즈 처리부(13)의 출력에 대응한 고역 정보가 포함되어 있고, 디코드 처리부(11)는, 그 고역 정보를 대역 확장부(41)에 공급한다.
- [0119] 또한, 렌더링 처리부(12)에서는, 디코드 처리부(11)로부터 공급된 오브젝트 위치 정보 및 오브젝트 신호에 기초하여 VBAP 등의 렌더링 처리가 행해지고, 그 결과 얻어진 가상 스피커 신호가 버추얼라이즈 처리부(13)에 공급된다.
- [0120] 버추얼라이즈 처리부(13)에서는, 버추얼라이즈 처리로서 HRTF 처리가 행해진다. 즉, 버추얼라이즈 처리부(13)에서는, 렌더링 처리부(12)로부터 공급된 가상 스피커 신호와, 사전에 부여된 전달 함수에 대응하는 HRTF 계수에 기초하는 컨벌루션 처리와, 그 결과 얻어지는 신호를 더하는 가산 처리가 HRTF 처리로서 행해진다. 버추얼라이즈 처리부(13)는, HRTF 처리에 의해 얻어진 오디오 신호를 대역 확장부(41)에 공급한다.
- [0121] 이 예에서는, 예를 들어 디코드 처리부(11)로부터 렌더링 처리부(12)로 공급되는 오브젝트 신호는, 샘플링 주파수가 48kHz인 저FS 오브젝트 신호로 된다.
- [0122] 그러한 경우, 렌더링 처리부(12)로부터 버추얼라이즈 처리부(13)로 공급되는 가상 스피커 신호도 샘플링 주파수가 48kHz인 신호로 되므로, 버추얼라이즈 처리부(13)로부터 대역 확장부(41)로 공급되는 오디오 신호의 샘플링 주파수도 48kHz로 된다.
- [0123] 이하에는, 버추얼라이즈 처리부(13)로부터 대역 확장부(41)로 공급되는 오디오 신호를, 특히 저FS 오디오 신호

라고도 칭하는 것으로 한다. 이러한 저FS 오디오 신호는, 오브젝트 신호에 대하여 렌더링 처리나 버추얼라이즈 처리 등의 신호 처리를 실시함으로써 얻어진, 헤드폰이나 실제 스피커 등의 재생 장치를 구동시켜 소리를 출력시키는 구동 신호이다.

- [0124] 대역 확장부(41)는, 디코드 처리부(11)로부터 공급된 고역 정보에 기초하여, 버추얼라이즈 처리부(13)로부터 공급된 저FS 오디오 신호에 대하여 대역 확장 처리를 행함으로써 출력 오디오 신호를 생성하여, 후단에 출력한다. 대역 확장부(41)에서 얻어지는 출력 오디오 신호는, 예를 들어 샘플링 주파수가 96kHz인 신호로 된다.
- [0125] 그런데, 버추얼라이즈 처리로서의 HRTF 처리에 사용되는 HRTF 계수는, 수청자인 유저 개인의 귀나 얼굴의 형상에 크게 의존하는 것이 잘 알려져 있다.
- [0126] 일반적인 가상 서라운드 대응의 헤드폰 등에서는, 유저 개인에 적합한 개인용 HRTF 계수를 취득하는 것이 곤란하기 때문에, 평균적인 귀나 얼굴의 형상을 위한 범용적인 HRTF 계수, 즉, 소위 범용 HRTF 계수가 많이 사용되고 있다.
- [0127] 그러나, 범용의 HRTF 계수를 사용한 경우, 개인용 HRTF 계수를 사용한 경우와 비교하여, 음원의 정위감이나 음질 그 자체가 크게 떨어지는 것이 알려져 있다.
- [0128] 그 때문에, 유저 개인에 적합한 HRTF 계수를 더 간단하게 취득하는 측정 방법 등도 제안되어 있고, 그러한 측정 방법은, 예를 들어 국제 공개 제2018/110269호 등에 상세하게 기재되어 있다.
- [0129] 이하에는, 사람의 평균적인 귀나 얼굴의 형상에 대하여 측정 또는 생성된, 범용적인 HRTF 계수를, 특히 범용 HRTF 계수라고도 칭하는 것으로 한다.
- [0130] 또한, 이하, 유저 개인에 대하여 측정 또는 생성된, 유저 개인의 귀나 얼굴의 형상에 대응하는 HRTF 계수, 즉 유저 개인마다의 HRTF 계수를, 특히 개인용 HRTF 계수라고도 칭하는 것으로 한다.
- [0131] 또한, 개인용 HRTF 계수는, 유저 개인에 대하여 측정 또는 생성된 것에 한정되지 않고, 귀나 얼굴의 형상마다 등에 측정 또는 생성된 복수의 HRTF 계수 중에서, 유저의 대략적인 귀나 얼굴의 형상이나 연령, 성별 등의 유저 개인에 관한 정보에 기초하여 선택된, 유저 개인에 적합한 HRTF 계수여도 된다.
- [0132] 이상과 같이, 유저마다, 그 유저에 적합한 HRTF 계수는 다르다.
- [0133] 예를 들어, 도 6에 나타난 신호 처리 장치(71)의 버추얼라이즈 처리부(13)에서, 개인용 HRTF 계수를 이용하는 것으로 하면, 대역 확장부(41)에서 이용되는 고역 정보도 개인용 HRTF 계수에 대응한 것으로 하는 것이 바람직하다.
- [0134] 그러나, 입력 비트 스트림에 포함되어 있는 고역 정보는, 범용 HRTF 계수를 사용하여 HRTF 처리를 행함으로써 얻어지는 오디오 신호에 대하여 대역 확장 처리를 행하는 것을 상정한 것인 범용 고역 정보로 되어 있다.
- [0135] 그 때문에, 개인용 HRTF 계수를 사용하여 HRTF 처리를 행함으로써 얻어지는 오디오 신호에 대하여, 입력 비트 스트림에 포함되어 있는 고역 정보를 그대로 사용하여 대역 확장 처리를 행하면, 얻어지는 출력 오디오 신호에 큰 음질 열화가 발생해 버리는 경우가 있다.
- [0136] 한편, 미리 개인용 HRTF 계수를 사용하는 것을 전제로 한, 유저마다, 즉 개인용 HRTF 계수마다 생성한 고역 정보(개인용 고역 정보)를 입력 비트 스트림에 저장하여 전송하는 것은 운용상, 용이하지 않다.
- [0137] 이것은, 오브젝트 오디오를 재생하는 유저(개인) 각각을 위해 입력 비트 스트림을 준비하거나, 개인용 HRTF 계수마다, 그것들의 개인용 HRTF 계수에 대응하는 개인용 고역 정보를 준비하거나 할 필요가 있기 때문이다. 또한, 그와 같이 하면, 오브젝트 오디오(입력 비트 스트림)를 배신하는 측, 즉 부호화측의 서버 등의 스토리지 용량도 압박되어 버린다.
- [0138] 그래서, 본 기술에서는, 범용 HRTF 계수를 전제로 한 범용 고역 정보와, 범용 HRTF 계수와, 개인용 HRTF 계수를 사용하여, 개인용 고역 정보를 재생 장치측(복호측)에서 생성하도록 했다.
- [0139] 이에 의해, 예를 들어 처리 부하가 높은 디코드 처리, 렌더링 처리, HRTF 처리를 낮은 샘플링 주파수에서 행하고, HRTF 처리 후의 최종적인 신호에 대하여, 생성된 개인용 고역 정보에 기초하는 대역 확장 처리를 행할 수 있게 된다. 따라서, 전체에 있어서의 처리 부하를 저감시켜, 저비용의 프로세서나 배터리에서도 고품질의 오디오 재생을 실현할 수 있다.

- [0140] <신호 처리 장치의 구성예>
- [0141] 도 7은, 본 기술을 적용한 신호 처리 장치(101)의 일 실시 형태의 구성예를 나타내는 도면이다. 또한, 도 7에 있어서 도 6에 있어서의 경우와 대응하는 부분에는 동일한 부호를 부여하고 있고, 그 설명은 적절히 생략한다.
- [0142] 신호 처리 장치(101)는, 예를 들어 스마트폰이나 퍼스널 컴퓨터 등을 포함하고, 디코드 처리부(11), 렌더링 처리부(12), 버추얼라이즈 처리부(13), 개인용 고역 정보 생성부(121), HRTF 계수 기록부(122) 및 대역 확장부(41)를 갖고 있다.
- [0143] 신호 처리 장치(101)의 구성은, 새롭게 개인용 고역 정보 생성부(121) 및 HRTF 계수 기록부(122)를 마련한 점에서 신호 처리 장치(71)의 구성과 달리, 그 밖의 점에서는 신호 처리 장치(71)와 동일한 구성으로 되어 있다.
- [0144] 디코드 처리부(11)는, 도시하지 않은 서버 등으로부터, 오브젝트 오디오가 부호화된 오브젝트 신호나, 오브젝트 위치 정보 등이 포함된 메타데이터, 범용 고역 정보 등이 포함되어 있는 입력 비트 스트림을 취득(수신)한다.
- [0145] 입력 비트 스트림에 포함되어 있는 범용 고역 정보는, 기본적으로는 신호 처리 장치(71)의 디코드 처리부(11)가 취득하는 입력 비트 스트림에 포함되어 있는 고역 정보와 동일한 것으로 되어 있다.
- [0146] 디코드 처리부(11)는, 수신하거나 하여 취득한 입력 비트 스트림을, 부호화된 오브젝트 신호나 메타데이터, 범용 고역 정보에 비다중화함과 함께, 부호화되어 있는 오브젝트 신호나 메타데이터를 디코드한다.
- [0147] 디코드 처리부(11)는, 입력 비트 스트림에 대한 비다중화 및 디코드 처리에 의해 얻어진 범용 고역 정보를 개인용 고역 정보 생성부(121)에 공급함과 함께, 오브젝트 위치 정보 및 오브젝트 신호를 렌더링 처리부(12)에 공급한다.
- [0148] 여기서, 입력 비트 스트림에는, 범용 HRTF 계수를 사용하여 버추얼라이즈 처리부(13)에서 HRTF 처리를 행했을 때의 버추얼라이즈 처리부(13)의 출력에 대응한 범용 고역 정보가 포함되어 있다. 즉, 범용 고역 정보는, 범용 HRTF 계수를 사용하여 HRTF 처리를 행함으로써 얻어지는 HRTF 출력 신호의 대역 확장을 위한 고역 정보이다.
- [0149] 렌더링 처리부(12)는, 디코드 처리부(11)로부터 공급된 오브젝트 위치 정보 및 오브젝트 신호에 기초하여 VBAP 등의 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 버추얼라이즈 처리부(13)에 공급한다.
- [0150] 버추얼라이즈 처리부(13)는, 렌더링 처리부(12)로부터 공급된 가상 스피커 신호와, HRTF 계수 기록부(122)로부터 공급된, 사전에 부여된 전달 함수에 대응하는 개인용 HRTF 계수에 기초하여, 버추얼라이즈 처리로서 HRTF 처리를 행하고, 그 결과 얻어진 오디오 신호를 대역 확장부(41)에 공급한다.
- [0151] 예를 들어 HRTF 처리에서는, 가상 스피커마다의 가상 스피커 신호와 개인용 HRTF 계수의 컨벌루션 처리 및 그것들의 가상 스피커마다의 컨벌루션 처리에 의해 얻어진 신호를 더하는 가산 처리가 행해진다.
- [0152] 또한, 이하, 버추얼라이즈 처리부(13)에서의 HRTF 처리에 의해 얻어진 오디오 신호를, 특히 HRTF 출력 신호라고도 칭하는 것으로 한다. HRTF 출력 신호는, 오브젝트 신호에 대하여 렌더링 처리나 버추얼라이즈 처리 등의 신호 처리를 실시함으로써 얻어진, 헤드폰 등의 재생 장치를 구동시켜 소리를 출력시키는 구동 신호이다.
- [0153] 신호 처리 장치(101)에서는, 예를 들어 디코드 처리부(11)로부터 렌더링 처리부(12)로 공급되는 오브젝트 신호는, 샘플링 주파수가 48kHz인 저FS 오브젝트 신호로 된다.
- [0154] 그러한 경우, 렌더링 처리부(12)로부터 버추얼라이즈 처리부(13)로 공급되는 가상 스피커 신호도 샘플링 주파수가 48kHz인 신호가 되므로, 버추얼라이즈 처리부(13)로부터 대역 확장부(41)로 공급되는 HRTF 출력 신호의 샘플링 주파수도 48kHz로 된다.
- [0155] 신호 처리 장치(101)에서는, 렌더링 처리부(12) 및 버추얼라이즈 처리부(13)가, 메타데이터(오브젝트 위치 정보)나 개인용 HRTF 계수, 오브젝트 신호에 기초하여 렌더링 처리나 버추얼라이즈 처리를 포함하는 신호 처리를 행하여, HRTF 출력 신호를 생성하는 신호 처리부로서 기능한다고 할 수 있다. 이 경우, 신호 처리에는, 적어도 버추얼라이즈 처리가 포함되어 있으면 된다.
- [0156] 개인용 고역 정보 생성부(121)는, 디코드 처리부(11)로부터 공급된 범용 고역 정보와, HRTF 계수 기록부(122)로부터 공급된 범용 HRTF 계수 및 개인용 HRTF 계수에 기초하여 개인용 고역 정보를 생성하여, 대역 확장부(41)에 공급한다.
- [0157] 이 개인용 고역 정보는, 개인용 HRTF 계수를 사용하여 HRTF 처리를 행함으로써 얻어지는 HRTF 출력 신호의 대역

확장을 위한 고역 정보이다.

- [0158] HRTF 계수 기록부(122)는, 미리 기록되어 있거나, 또는 필요에 따라 외부의 장치로부터 취득한 범용 HRTF 계수나 개인용 HRTF 계수를 기록(유지)한다.
- [0159] HRTF 계수 기록부(122)는, 기록되어 있는 개인용 HRTF 계수를 버추얼라이즈 처리부(13)에 공급하거나, 기록되어 있는 범용 HRTF 계수 및 개인용 HRTF 계수를 개인용 고역 정보 생성부(121)에 공급하거나 한다.
- [0160] 일반적으로, 범용 HRTF 계수는 미리 재생 장치의 기록 영역 내에 저장되어 있기 때문에, 이 예에서도 범용 HRTF 계수가 재생 장치로서 기능하는 신호 처리 장치(101)의 HRTF 계수 기록부(122)에 미리 기록되어 있도록 할 수 있다.
- [0161] 또한, 개인용 HRTF 계수는, 네트워크상의 서버 등으로부터 취득되도록 할 수 있다.
- [0162] 그러한 경우, 예를 들어 재생 장치로서 기능하는 신호 처리 장치(101) 자신이나, 신호 처리 장치(101)에 접속된, 스마트폰 등의 단말 장치에 의해, 유저의 얼굴 화상이나 귀 화상 등의 화상 데이터가 촬영에 의해 생성된다.
- [0163] 그리고, 신호 처리 장치(101)는, 유저에 대하여 얻어진 화상 데이터를 서버로 송신하고, 서버는 신호 처리 장치(101)로부터 수신한 화상 데이터에 기초하여, 유지하고 있는 HRTF 계수에 대한 변환 처리를 행하여 유저 개인의 개인용 HRTF 계수를 생성하고, 신호 처리 장치(101)로 송신한다. HRTF 계수 기록부(122)는, 이와 같이 하여 서버로부터 송신되고, 신호 처리 장치(101)에 의해 수신된 개인용 HRTF 계수를 취득하여, 기록한다.
- [0164] 대역 확장부(41)는, 개인용 고역 정보 생성부(121)로부터 공급된 개인용 고역 정보에 기초하여, 버추얼라이즈 처리부(13)로부터 공급된 HRTF 출력 신호에 대하여 대역 확장 처리를 행함으로써 출력 오디오 신호를 생성하여, 후단에 출력한다. 대역 확장부(41)에서 얻어지는 출력 오디오 신호는, 예를 들어 샘플링 주파수가 96kHz인 신호로 된다.
- [0165] <개인용 고역 정보 생성부의 구성예>
- [0166] 상술한 바와 같이, 개인용 고역 정보 생성부(121)에서는, 범용 고역 정보와, 범용 HRTF 계수와, 개인용 HRTF 계수에 기초하여 개인용 고역 정보가 생성된다.
- [0167] 원래라면, 입력 비트 스트림 내에 개인용 고역 정보를 다중화해야 하지만, 그렇게 하면, 서버상에서 각 유저의 개인용의 입력 비트 스트림을 유지하게 되어, 서버의 스토리지 용량의 관점에서 바람직하지 않다.
- [0168] 따라서 본 기술에서는, 입력 비트 스트림 내에는 범용 고역 정보가 다중화되고, 어떤 방법에 의해 개인용 고역 정보 생성부(121)에 의해 취득된 개인용 HRTF 계수와 범용 HRTF 계수가 사용되어, 개인용 고역 정보가 생성된다.
- [0169] 개인용 고역 정보 생성부(121)에서의 개인용 고역 정보의 생성은, 어떤 방법에 의해 실현되어도 되지만, 일례로서, 예를 들어 DNN(Deep Neural Network)과 같은 심층 학습 기술을 사용하여 실현할 수 있다.
- [0170] 여기서, 개인용 고역 정보 생성부(121)가 DNN에 의해 구성되는 경우를 예로서 설명한다.
- [0171] 예를 들어 개인용 고역 정보 생성부(121)는, 미리 기계 학습에 의해 생성된 DNN을 구성하는 계수와, DNN의 입력이 되는 범용 고역 정보, 범용 HRTF 계수 및 개인용 HRTF 계수에 기초하는, DNN(뉴럴 네트워크)에 의한 연산을 행함으로써, 개인용 고역 정보를 생성한다.
- [0172] 그러한 경우, 개인용 고역 정보 생성부(121)는, 예를 들어 도 8에 나타낸 바와 같이 구성된다.
- [0173] 개인용 고역 정보 생성부(121)는, MLP(Multi-Layer Perceptron)(151), MLP(152), RNN(Recurrent Neural Network)(153), 특징량 합성부(154) 및 MLP(155)를 갖고 있다.
- [0174] MLP(151)는, 비선형적으로 활성화되는 노드의 3개 이상의 층, 즉 입력층과 출력층과 1개 이상의 은닉층으로 구성되는 MLP이다. MLP는 DNN에 있어서 일반적으로 사용되는 기술 중 하나이다.
- [0175] MLP(151)는, HRTF 계수 기록부(122)로부터 공급된 범용 HRTF 계수를 MLP의 입력으로 하는 벡터 gh\_in으로 하고, 벡터 gh\_in에 기초하는 연산 처리를 행함으로써, 범용 HRTF 계수의 어떤 특징을 나타내는 데이터인 벡터 gh\_out을 생성(산출)하여, 특징량 합성부(154)에 공급한다.
- [0176] 또한, MLP의 입력이 되는 벡터 gh\_in은, 범용 HRTF 계수 그 자체여도 되고, 후단의 계산 자원을 삭감하기 위해,

범용 HRTF 계수에 대하여 어떤 전처리를 행함으로써 얻어지는 특징량이어도 된다.

- [0177] MLP(152)는, MLP(151)와 마찬가지로 MLP를 포함하고, HRTF 계수 기록부(122)로부터 공급된 개인용 HRTF 계수를 MLP의 입력으로 하는 벡터 `ph_in`으로 하고, 벡터 `ph_in`에 기초하는 연산 처리를 행함으로써, 개인용 HRTF 계수의 어떤 특징을 나타내는 데이터인 벡터 `ph_out`을 생성하여, 특징량 합성부(154)에 공급한다.
- [0178] 또한, 벡터 `ph_in`도 개인용 HRTF 계수 그 자체여도 되고, 어떤 전처리를 개인용 HRTF 계수에 대하여 실시함으로써 얻어지는 특징량이어도 된다.
- [0179] RNN(153)은, 예를 들어 일반적으로 입력층, 은닉층, 출력층의 3층으로 구성되는 RNN을 포함한다. 이 RNN에서는, 예를 들어 은닉층의 출력이 은닉층의 입력에 피드백되도록 되어 있고, RNN은, 시계열의 데이터에 대하여 적합한 뉴럴 네트워크의 구조로 되어 있다.
- [0180] 또한, 여기서는 개인용 고역 정보의 생성에 RNN을 사용하는 예에 대하여 설명하지만, 본 기술은, 개인용 고역 정보 생성부(121)로서의 DNN의 구성에 의존하는 것은 아니고, RNN 대신에, 예를 들어 더 장기의 시계열 데이터에 적합한 뉴럴 네트워크의 구조인 LSTM(Long Short Term Memory) 등을 사용하도록 해도 된다.
- [0181] RNN(153)은, 디코드 처리부(11)로부터 공급된 범용 고역 정보를 입력으로 하는 벡터 `ge_in(n)`으로 하고, 벡터 `ge_in(n)`에 기초하는 연산 처리를 행함으로써, 범용 고역 정보의 어떤 특징을 나타내는 데이터인 벡터 `ge_out(n)`을 생성(산출)하여, 특징량 합성부(154)에 공급한다.
- [0182] 또한, 벡터 `ge_in(n)` 및 벡터 `ge_out(n)`에 있어서의 `n`은, 오브젝트 신호의 시간 프레임의 인덱스를 나타내고 있다. 특히 RNN(153)에서는, 1프레임분의 개인용 고역 정보를 생성하기 위해, 복수 프레임분의 벡터 `ge_in(n)`이 사용된다.
- [0183] 특징량 합성부(154)는, MLP(151)로부터 공급된 벡터 `gh_out`, MLP(152)로부터 공급된 벡터 `ph_out` 및 RNN(153)으로부터 공급된 벡터 `ge_out(n)`을 벡터 연결함으로써, 하나의 벡터 `co_out(n)`을 생성하여, MLP(155)에 공급한다.
- [0184] 또한, 여기서는 특징량 합성부(154)에 있어서의 특징량 합성의 방법으로서 벡터 연결을 사용하고 있지만, 이에 한정되지는 않고, 다른 어떤 방법에 의해 벡터 `co_out(n)`을 생성해도 된다. 예를 들어 특징량 합성부(154)에 있어서, `max-pooling`이라고 불리는 방법에 의해 특징량 합성을 행하여, 특징을 충분히 표현할 수 있는 콤팩트한 사이즈로 벡터가 합성되도록 해도 된다.
- [0185] MLP(155)는, 예를 들어 입력층, 출력층 및 1 이상의 은닉층을 갖는 MLP를 포함하고, 특징량 합성부(154)로부터 공급된 벡터 `co_out(n)`에 기초하여 연산 처리를 행하고, 그 결과 얻어진 벡터 `pe_out(n)`을 개인용 고역 정보로서 대역 확장부(41)에 공급한다.
- [0186] 이상과 같은 개인용 고역 정보 생성부(121)로서 기능하는 DNN을 구성하는 MLP(151), MLP(152), RNN(153), MLP(155) 등의 MLP나 RNN을 구성하는 계수는, 사전에 교사 데이터를 사용하여 기계 학습을 행함으로써 얻을 수 있다.
- [0187] <입력 비트 스트림의 선택스에>
- [0188] 신호 처리 장치(101)에서는, 개인용 고역 정보의 생성을 위해 범용 고역 정보가 필요하고, 입력 비트 스트림에는, 범용 고역 정보가 저장되어 있다.
- [0189] 여기서, 디코드 처리부(11)에 공급되는 입력 비트 스트림의 선택스에, 즉 입력 비트 스트림의 포맷예를 도 9에 나타낸다.
- [0190] 도 9에 있어서 「`num_objects`」는 오브젝트의 총수를 나타내고 있고, 「`object_compressed_data`」는 부호화(압축)된 오브젝트 신호를 나타내고 있다.
- [0191] 또한, 「`position_azimuth`」는 오브젝트의 구면 좌표계에 있어서의 수평 각도를 나타내고 있고, 「`position_elevation`」은 오브젝트의 구면 좌표계에 있어서의 수직 각도를 나타내고 있고, 「`position_radius`」는 구면 좌표계 원점으로부터 오브젝트까지의 거리(반경)를 나타내고 있다. 여기서는, 이것들의 수평 각도, 수직 각도 및 거리를 포함하는 정보가 오브젝트의 위치를 나타내는 오브젝트 위치 정보로 되어 있다.
- [0192] 따라서, 이 예에서는, 「`num_objects`」에 의해 나타나는 오브젝트수만큼, 부호화된 오브젝트 신호 및 오브젝트 위치 정보가 입력 비트 스트림에 포함되어 있다.

- [0193] 또한, 도 9에 있어서 「num\_output」은, 출력 채널수, 즉 HRTF 출력 신호의 채널수를 나타내고 있고, 「output\_bwe\_data」는 범용 고역 정보를 나타내고 있다. 따라서, 이 예에서는, HRTF 출력 신호의 채널마다 범용 고역 정보가 저장되어 있다.
- [0194] <신호 생성 처리의 설명>
- [0195] 이어서, 신호 처리 장치(101)의 동작에 대하여 설명한다. 즉, 이하, 도 10의 흐름도를 참조하여, 신호 처리 장치(101)에 의한 신호 생성 처리에 대하여 설명한다.
- [0196] 스텝 S11에 있어서 디코드 처리부(11)는, 공급된 입력 비트 스트림에 대하여 비다중화 및 디코드 처리를 행하고, 그 결과 얻어진 범용 고역 정보를 개인용 고역 정보 생성부(121)에 공급함과 함께, 오브젝트 위치 정보 및 오브젝트 신호를 렌더링 처리부(12)에 공급한다.
- [0197] 여기서는, 예를 들어 도 9에 나타난 「output\_bwe\_data」에 의해 나타나는 범용 고역 정보가 입력 비트 스트림으로부터 추출되어, 개인용 고역 정보 생성부(121)로 공급된다.
- [0198] 스텝 S12에 있어서 렌더링 처리부(12)는, 디코드 처리부(11)로부터 공급된 오브젝트 위치 정보 및 오브젝트 신호에 기초하여 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 버추얼라이즈 처리부(13)에 공급한다. 예를 들어 스텝 S12에서는, VBAP 등이 렌더링 처리로서 행해진다.
- [0199] 스텝 S13에 있어서 버추얼라이즈 처리부(13)는 버추얼라이즈 처리를 행한다. 예를 들어 스텝 S13에서는, HRTF 처리가 버추얼라이즈 처리로서 행해진다.
- [0200] 이 경우, 버추얼라이즈 처리부(13)는, 렌더링 처리부(12)로부터 공급된 각 가상 스피커의 가상 스피커 신호와, HRTF 계수 기록부(122)로부터 공급된 채널마다의 각 가상 스피커의 개인용 HRTF 계수를 컨볼루션하고, 그 결과 얻어진 신호를 채널마다 가산하는 처리를 HRTF 처리로서 행한다. 버추얼라이즈 처리부(13)는, HRTF 처리에 의해 얻어진 HRTF 출력 신호를 대역 확장부(41)에 공급한다.
- [0201] 스텝 S14에 있어서 개인용 고역 정보 생성부(121)는, 디코드 처리부(11)로부터 공급된 범용 고역 정보와, HRTF 계수 기록부(122)로부터 공급된 범용 HRTF 계수 및 개인용 HRTF 계수에 기초하여 개인용 고역 정보를 생성하여, 대역 확장부(41)에 공급한다.
- [0202] 예를 들어 스텝 S14에서는, DNN을 구성하는 개인용 고역 정보 생성부(121)의 MLP(151) 내지 MLP(155)에 의해, 개인용 고역 정보가 생성된다.
- [0203] 구체적으로는, MLP(151)는, HRTF 계수 기록부(122)로부터 공급된 범용 HRTF 계수, 즉 벡터 gh\_in에 기초하여 연산 처리를 행하고, 그 결과 얻어진 벡터 gh\_out을 특징량 합성부(154)에 공급한다.
- [0204] MLP(152)는, HRTF 계수 기록부(122)로부터 공급된 개인용 HRTF 계수, 즉 벡터 ph\_in에 기초하여 연산 처리를 행하고, 그 결과 얻어진 벡터 ph\_out을 특징량 합성부(154)에 공급한다.
- [0205] RNN(153)은, 디코드 처리부(11)로부터 공급된 범용 고역 정보, 즉 벡터 ge\_in(n)에 기초하여 연산 처리를 행하고, 그 결과 얻어진 벡터 ge\_out(n)을 특징량 합성부(154)에 공급한다.
- [0206] 또한, 특징량 합성부(154)는, MLP(151)로부터 공급된 벡터 gh\_out, MLP(152)로부터 공급된 벡터 ph\_out 및 RNN(153)으로부터 공급된 벡터 ge\_out(n)을 벡터 연결하고, 그 결과 얻어진 벡터 co\_out(n)을 MLP(155)에 공급한다.
- [0207] MLP(155)는, 특징량 합성부(154)로부터 공급된 벡터 co\_out(n)에 기초하여 연산 처리를 행하고, 그 결과 얻어진 벡터 pe\_out(n)을 개인용 고역 정보로서 대역 확장부(41)에 공급한다.
- [0208] 스텝 S15에 있어서 대역 확장부(41)는, 개인용 고역 정보 생성부(121)로부터 공급된 개인용 고역 정보에 기초하여, 버추얼라이즈 처리부(13)로부터 공급된 HRTF 출력 신호에 대하여 대역 확장 처리를 행하고, 그 결과 얻어진 출력 오디오 신호를 후단에 출력한다. 이와 같이 하여 출력 오디오 신호가 생성되면, 신호 생성 처리는 종료된다.
- [0209] 이상과 같이 하여 신호 처리 장치(101)는, 입력 비트 스트림으로부터 추출된(읽어내어진) 범용 고역 정보를 사용하여 개인용 고역 정보를 생성하고, 그 개인용 고역 정보를 사용하여 대역 확장 처리를 행하여 출력 오디오 신호를 생성한다.

- [0210] 이 경우, 렌더링 처리나 HRTF 처리가 행해져 얻어진, 낮은 샘플링 주파수의 HRTF 출력 신호에 대하여 대역 확장 처리를 행하도록 함으로써, 신호 처리 장치(101)에 있어서의 처리 부하, 즉 연산량을 저감시킬 수 있다.
- [0211] 또한, HRTF 처리에 사용되는 개인용 HRTF 계수에 대응하는 개인용 고역 정보를 생성하여 대역 확장 처리를 행함으로써, 고품질의 출력 오디오 신호를 얻을 수 있다.
- [0212] 따라서, 신호 처리 장치(101)가 저비용의 장치라도 고품질의 오디오 재생을 행할 수 있다.
- [0213] <학습 장치의 구성예>
- [0214] 이어서, 개인용 고역 정보 생성부(121)로서의 DNN(뉴럴 네트워크)을 구성하는 계수, 즉 MLP(151), MLP(152), RNN(153) 및 MLP(155)를 구성하는 계수를, 개인용 고역 정보 생성 계수 데이터로서 생성하는 학습 장치에 대하여 설명한다.
- [0215] 그러한 학습 장치는, 예를 들어 도 11에 나타낸 바와 같이 구성된다.
- [0216] 학습 장치(201)는, 렌더링 처리부(211), 개인용 HRTF 처리부(212), 개인용 고역 정보 계산부(213), 범용 HRTF 처리부(214), 범용 고역 정보 계산부(215) 및 개인용 고역 정보 학습부(216)를 갖고 있다.
- [0217] 렌더링 처리부(211)는, 공급된 오브젝트 위치 정보와 오브젝트 신호에 기초하여, 렌더링 처리부(12)에 있어서의 경우와 마찬가지로 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 개인용 HRTF 처리부(212) 및 범용 HRTF 처리부(214)에 공급한다.
- [0218] 또한, 렌더링 처리부(211)의 후단에 있어서, 교사 데이터로서 개인용 고역 정보가 필요해지기 때문에, 렌더링 처리부(211)의 출력인 가상 스피커 신호, 즉 렌더링 처리부(211)의 입력이 되는 오브젝트 신호에는 고역 정보가 포함되어 있을 필요가 있다.
- [0219] 예를 들어, 신호 처리 장치(101)의 버추얼라이즈 처리부(13)의 출력인 HRTF 출력 신호가 샘플링 주파수 48kHz의 신호라고 하면, 렌더링 처리부(211)에 입력되는 오브젝트 신호의 샘플링 주파수는 96kHz 등으로 된다.
- [0220] 이 경우, 렌더링 처리부(211)에서는, 샘플링 주파수 96kHz에서 VBAP 등의 렌더링 처리가 행해져, 샘플링 주파수가 96kHz인 가상 스피커 신호가 생성된다.
- [0221] 또한, 이하에 있어서는, 버추얼라이즈 처리부(13)의 출력인 HRTF 출력 신호가 샘플링 주파수 48kHz의 신호인 것으로 하여 설명을 행하지만, 본 기술에서는, 각 신호의 샘플링 주파수는, 이 예에 제한되는 것은 아니다. 예를 들어 HRTF 출력 신호의 샘플링 주파수가 44.1kHz이고, 렌더링 처리부(211)에 입력되는 오브젝트 신호의 샘플링 주파수가 88.2kHz여도 된다.
- [0222] 개인용 HRTF 처리부(212)는, 공급된 개인용 HRTF 계수와, 렌더링 처리부(211)로부터 공급된 가상 스피커 신호에 기초하여 HRTF 처리(이하, 특히 개인용 HRTF 처리라고도 칭함)를 행하고, 그 결과 얻어진 개인용 HRTF 출력 신호를 개인용 고역 정보 계산부(213)에 공급한다. 개인용 HRTF 처리에서 얻어지는 개인용 HRTF 출력 신호는, 샘플링 주파수가 96kHz인 신호이다.
- [0223] 이 예에서는, 렌더링 처리부(211) 및 개인용 HRTF 처리부(212)가, 메타데이터(오브젝트 위치 정보)나 개인용 HRTF 계수, 오브젝트 신호에 기초하여 렌더링 처리나 버추얼라이즈 처리(개인용 HRTF 처리)를 포함하는 신호 처리를 행하여, 개인용 HRTF 출력 신호를 생성하는 하나의 신호 처리부로서 기능한다고 할 수 있다. 이 경우, 신호 처리에는, 적어도 버추얼라이즈 처리가 포함되어 있으면 된다.
- [0224] 개인용 고역 정보 계산부(213)는, 개인용 HRTF 처리부(212)로부터 공급된 개인용 HRTF 출력 신호에 기초하여 개인용 고역 정보를 생성(계산)하고, 얻어진 개인용 고역 정보를 학습 시의 교사 데이터로서 개인용 고역 정보 학습부(216)에 공급한다.
- [0225] 예를 들어 개인용 고역 정보 계산부(213)에서는, 도 5를 참조하여 설명한 바와 같이, 개인용 HRTF 출력 신호의 각 고역 서브 밴드의 평균 진폭값이 개인용 고역 정보로서 요구된다.
- [0226] 즉, 샘플링 주파수가 96kHz인 개인용 HRTF 출력 신호에 대하여, 대역 통과 필터 बैं크를 적용하여 각 고역 서브 밴드의 고역 서브 밴드 신호를 생성한 후, 고역 서브 밴드 신호의 시간 프레임의 평균 진폭값을 계산함으로써 개인용 고역 정보를 얻을 수 있다.
- [0227] 범용 HRTF 처리부(214)는, 공급된 범용 HRTF 계수와, 렌더링 처리부(211)로부터 공급된 가상 스피커 신호에 기

초하여 HRTF 처리(이하, 특히 범용 HRTF 처리라고도 칭함)를 행하고, 그 결과 얻어진 범용 HRTF 출력 신호를 범용 고역 정보 계산부(215)에 공급한다. 범용 HRTF 출력 신호는, 샘플링 주파수가 96kHz인 신호이다.

- [0228] 이 예에서는, 렌더링 처리부(211) 및 범용 HRTF 처리부(214)가, 메타데이터(오브젝트 위치 정보)나 범용 HRTF 계수, 오브젝트 신호에 기초하여 렌더링 처리나 버추얼라이즈 처리(범용 HRTF 처리)를 포함하는 신호 처리를 행하여, 범용 HRTF 출력 신호를 생성하는 하나의 신호 처리부로서 기능한다고 할 수 있다. 이 경우, 신호 처리에는, 적어도 버추얼라이즈 처리가 포함되어 있으면 된다.
- [0229] 범용 고역 정보 계산부(215)는, 범용 HRTF 처리부(214)로부터 공급된 범용 HRTF 출력 신호에 기초하여 범용 고역 정보를 생성(계산)하여, 개인용 고역 정보 학습부(216)에 공급한다. 범용 고역 정보 계산부(215)에서는, 개인용 고역 정보 계산부(213)에 있어서의 경우와 마찬가지로의 계산이 행해져, 범용 고역 정보가 생성된다.
- [0230] 입력 비트 스트림에는, 범용 고역 정보 계산부(215)에서 얻어지는 범용 고역 정보와 마찬가지로의 것이 도 9에 나타난 「output\_bwe\_data」로서 포함되어 있다.
- [0231] 또한, 범용 HRTF 처리부(214) 및 범용 고역 정보 계산부(215)에 있어서 행해지는 처리는, 개인용 HRTF 처리부(212) 및 개인용 고역 정보 계산부(213)에 있어서 행해지는 처리와 쌍을 이루는 것이고, 이들 처리는 기본적으로는 동일한 처리이다.
- [0232] 이들 처리의 차이는, 개인용 HRTF 처리부(212)의 입력이 개인용 HRTF 계수인 것에 비해, 범용 HRTF 처리부(214)의 입력이 범용 HRTF 계수인 점뿐이다. 즉, 입력되는 HRTF 계수만이 다르게 되어 있다.
- [0233] 개인용 고역 정보 학습부(216)는, 공급된 범용 HRTF 계수 및 개인용 HRTF 계수와, 개인용 고역 정보 계산부(213)로부터 공급된 개인용 고역 정보와, 범용 고역 정보 계산부(215)로부터 공급된 범용 고역 정보에 기초하여 학습(기계 학습)을 행하고, 그 결과 얻어진 개인용 고역 정보 생성 계수 데이터를 출력한다.
- [0234] 특히, 개인용 고역 정보 학습부(216)에서는, 개인용 고역 정보를 교사 데이터로 하는 기계 학습이 행해지고, 범용 HRTF 계수, 개인용 HRTF 계수 및 범용 고역 정보로부터, 개인용 고역 정보를 생성하기 위한 개인용 고역 정보 생성 계수 데이터가 생성된다.
- [0235] 이와 같이 하여 얻어진 개인용 고역 정보 생성 계수 데이터를 구성하는 각 계수를, 도 8의 개인용 고역 정보 생성부(121)의 MLP(151), MLP(152), RNN(153) 및 MLP(155)에서 사용하면, 학습 결과에 기초하는 개인용 고역 정보의 생성이 가능해진다.
- [0236] 예를 들어 개인용 고역 정보 학습부(216)에서 행해지는 학습 처리는, 개인용 고역 정보 생성부(121)에서의 처리 결과로서 출력되는 벡터  $pe\_out(n)$ 과, 교사 데이터로서의 개인용 고역 정보인 벡터  $tpe\_out(n)$ 의 오차를 평가함으로써 행해진다. 즉, 벡터  $pe\_out(n)$ 과 벡터  $tpe\_out(n)$ 의 오차가 최소가 되도록 학습이 행해진다.
- [0237] DNN을 구성하는 MLP(151) 등의 각 요소의 가중 계수의 초깃값은 랜덤한 것이 일반적이고, 오차 평가에 따라 각 계수를 조정하는 방법에 대해서도 BPTT(Back Propagation Through Time)와 같은 오차 역전파법에 기초하는 다양한 방법을 적용 할 수 있다.
- [0238] <학습 처리의 설명>
- [0239] 계속해서, 학습 장치(201)의 동작에 대하여 설명한다. 즉, 이하, 도 12의 흐름도를 참조하여, 학습 장치(201)에 의한 학습 처리에 대하여 설명한다.
- [0240] 스텝 S41에 있어서 렌더링 처리부(211)는, 공급된 오브젝트 위치 정보와 오브젝트 신호에 기초하여 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 개인용 HRTF 처리부(212) 및 범용 HRTF 처리부(214)에 공급한다.
- [0241] 스텝 S42에 있어서 개인용 HRTF 처리부(212)는, 공급된 개인용 HRTF 계수와, 렌더링 처리부(211)로부터 공급된 가상 스피커 신호에 기초하여 개인용 HRTF 처리를 행하고, 그 결과 얻어진 개인용 HRTF 출력 신호를 개인용 고역 정보 계산부(213)에 공급한다.
- [0242] 스텝 S43에 있어서 개인용 고역 정보 계산부(213)는, 개인용 HRTF 처리부(212)로부터 공급된 개인용 HRTF 출력 신호에 기초하여 개인용 고역 정보를 계산하고, 얻어진 개인용 고역 정보를 교사 데이터로서 개인용 고역 정보 학습부(216)에 공급한다.
- [0243] 스텝 S44에 있어서 범용 HRTF 처리부(214)는, 공급된 범용 HRTF 계수와, 렌더링 처리부(211)로부터 공급된 가상

스피커 신호에 기초하여 범용 HRTF 처리를 행하고, 그 결과 얻어진 범용 HRTF 출력 신호를 범용 고역 정보 계산부(215)에 공급한다.

- [0244] 스텝 S45에 있어서 범용 고역 정보 계산부(215)는, 범용 HRTF 처리부(214)로부터 공급된 범용 HRTF 출력 신호에 기초하여 범용 고역 정보를 계산하고, 개인용 고역 정보 학습부(216)에 공급한다.
- [0245] 스텝 S46에 있어서 개인용 고역 정보 학습부(216)는, 공급된 범용 HRTF 계수 및 개인용 HRTF 계수와, 개인용 고역 정보 계산부(213)로부터 공급된 개인용 고역 정보와, 범용 고역 정보 계산부(215)로부터 공급된 범용 고역 정보에 기초하여 학습을 행하여, 개인용 고역 정보 생성 계수 데이터를 생성한다.
- [0246] 학습 시에는, 범용 고역 정보, 범용 HRTF 계수 및 개인용 HRTF 계수를 입력으로 하고, 교사 데이터인 개인용 고역 정보를 출력으로 하는 DNN을 실현하기 위한 개인용 고역 정보 생성 계수 데이터가 생성된다. 이와 같이 하여 개인용 고역 정보 생성 계수 데이터가 생성되면, 학습 처리는 종료된다.
- [0247] 이상과 같이 하여 학습 장치(201)는, 범용 HRTF 계수나 개인용 HRTF 계수, 오브젝트 신호에 기초하여 학습을 행하여, 개인용 고역 정보 생성 계수 데이터를 생성한다.
- [0248] 이렇게 함으로써, 개인용 고역 정보 생성부(121)에서는, 입력된 범용 고역 정보, 범용 HRTF 계수 및 개인용 HRTF 계수로부터, 개인용 HRTF 계수에 대응하는 적절한 개인용 고역 정보를 예측에 의해 얻을 수 있게 된다.
- [0249] <인코더의 구성예>
- [0250] 계속해서, 도 9에 나타난 포맷의 입력 비트 스트림을 생성하는 인코더(부호화 장치)에 대하여 설명한다. 그러한 인코더는, 예를 들어 도 13에 나타난 바와 같이 구성된다.
- [0251] 도 13에 나타내는 인코더(301)는, 오브젝트 위치 정보 부호화부(311), 다운 샘플러(312), 오브젝트 신호 부호화부(313), 렌더링 처리부(314), 범용 HRTF 처리부(315), 범용 고역 정보 계산부(316) 및 다중화부(317)를 갖고 있다.
- [0252] 인코더(301)에는, 부호화 대상이 되는 오브젝트의 오브젝트 신호와, 그 오브젝트의 위치를 나타내는 오브젝트 위치 정보가 입력(공급)된다.
- [0253] 여기서는, 인코더(301)에 입력되는 오브젝트 신호는, 예를 들어 샘플링 주파수가 96kHz의 신호(FS96K 오브젝트 신호)인 것으로 한다.
- [0254] 오브젝트 위치 정보 부호화부(311)는, 입력된 오브젝트 위치 정보를 부호화하여, 다중화부(317)에 공급한다.
- [0255] 이에 의해, 부호화된 오브젝트 위치 정보로서, 예를 들어 도 9에 나타난 수평 각도 「position\_azimuth」, 수직 각도 「position\_elevation」 및 반경 「position\_radius」을 포함하는 부호화된 오브젝트 위치 정보(오브젝트 위치 데이터)가 얻어진다.
- [0256] 다운 샘플러(312)는, 입력된 샘플링 주파수가 96kHz인 오브젝트 신호에 대하여 다운 샘플링 처리, 즉 대역 제한을 행하고, 그 결과 얻어진 샘플링 주파수가 48kHz인 오브젝트 신호(FS48K 오브젝트 신호)를 오브젝트 신호 부호화부(313)에 공급한다.
- [0257] 오브젝트 신호 부호화부(313)는, 다운 샘플러(312)로부터 공급된 48kHz의 오브젝트 신호를 부호화하여 다중화부(317)에 공급한다. 이에 의해, 예를 들어 도 9에 나타난 「object\_compressed\_data」가 부호화된 오브젝트 신호로서 얻어진다.
- [0258] 또한, 오브젝트 신호 부호화부(313)에서의 부호화 방식은, MPEG-H Part 3: 3D audio 규격의 부호화 방식이어도 되고, 그밖의 부호화 방식이어도 된다. 즉, 오브젝트 신호 부호화부(313)에서의 부호화 방식과 디코드 처리부(11)에서의 복호 방식이 대응하는 것(동일 규격의 것)이면 된다.
- [0259] 렌더링 처리부(314)는, 입력된 오브젝트 위치 정보 및 96kHz의 오브젝트 신호에 기초하여 VBAP 등의 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 범용 HRTF 처리부(315)에 공급한다.
- [0260] 또한, 렌더링 처리부(314)에서의 렌더링 처리는, 복호측(재생측)인 신호 처리 장치(101)의 렌더링 처리부(12)에 있어서의 경우와 동일한 처리라면, VBAP에 한정되지는 않고 다른 어떤 렌더링 처리여도 된다.
- [0261] 범용 HRTF 처리부(315)는, 렌더링 처리부(314)로부터 공급된 가상 스피커 신호에 대하여, 범용 HRTF 계수를 사용한 HRTF 처리를 행하고, 그 결과 얻어진 96kHz의 범용 HRTF 출력 신호를 범용 고역 정보 계산부(316)에 공급한다.

다.

- [0262] 범용 HRTF 처리부(315)에서는, 도 11의 범용 HRTF 처리부(214)에 있어서의 범용 HRTF 처리와 마찬가지로 처리가 행해진다.
- [0263] 범용 고역 정보 계산부(316)는, 범용 HRTF 처리부(315)로부터 공급된 범용 HRTF 출력 신호에 기초하여 범용 고역 정보를 계산함과 함께, 얻어진 범용 고역 정보를 압축 부호화하여, 다중화부(317)에 공급한다.
- [0264] 범용 고역 정보 계산부(316)에서 생성되는 범용 고역 정보는, 예를 들어 도 5에 나타낸 각 고역 서브 밴드의 평균 진폭 정보(평균 진폭값)이다.
- [0265] 예를 들어 범용 고역 정보 계산부(316)는, 입력된 96kHz의 범용 HRTF 출력 신호에 대하여 대역 통과 필터뱅크에 기초하는 필터링을 행하여, 각 고역 서브 밴드의 고역 서브 밴드 신호를 얻는다. 그리고, 범용 고역 정보 계산부(316)는, 그것들의 각 고역 서브 밴드 신호의 시간 프레임의 평균 진폭값을 계산함으로써, 범용 고역 정보를 생성한다.
- [0266] 이에 의해, 예를 들어 도 9에 나타낸 「output\_bwe\_data」가 부호화된 범용 고역 정보로서 얻어진다.
- [0267] 다중화부(317)는, 오브젝트 위치 정보 부호화부(311)로부터 공급된 부호화된 오브젝트 위치 정보, 오브젝트 신호 부호화부(313)로부터 공급된 부호화된 오브젝트 신호 및 범용 고역 정보 계산부(316)로부터 공급된 부호화된 범용 고역 정보를 다중화한다.
- [0268] 다중화부(317)는, 오브젝트 위치 정보나 오브젝트 신호, 범용 고역 정보를 다중화하여 얻어진 출력 비트 스트림을 출력한다. 이 출력 비트 스트림은, 입력 비트 스트림으로서 신호 처리 장치(101)에 입력된다.
- [0269] <부호화 처리의 설명>
- [0270] 이어서, 인코더(301)의 동작에 대하여 설명한다. 즉, 이하, 도 14의 흐름도를 참조하여, 인코더(301)에 의한 부호화 처리에 대하여 설명한다.
- [0271] 스텝 S71에 있어서 오브젝트 위치 정보 부호화부(311)는, 입력된 오브젝트 위치 정보를 부호화하여, 다중화부(317)에 공급한다.
- [0272] 스텝 S72에 있어서 다운 샘플러(312)는, 입력된 오브젝트 신호를 다운 샘플링하여 오브젝트 신호 부호화부(313)에 공급한다.
- [0273] 스텝 S73에 있어서 오브젝트 신호 부호화부(313)는, 다운 샘플러(312)로부터 공급된 오브젝트 신호를 부호화하여 다중화부(317)에 공급한다.
- [0274] 스텝 S74에 있어서 렌더링 처리부(314)는, 입력된 오브젝트 위치 정보 및 오브젝트 신호에 기초하여 렌더링 처리를 행하고, 그 결과 얻어진 가상 스피커 신호를 범용 HRTF 처리부(315)에 공급한다.
- [0275] 스텝 S75에 있어서 범용 HRTF 처리부(315)는, 렌더링 처리부(314)로부터 공급된 가상 스피커 신호에 대하여, 범용 HRTF 계수를 사용한 HRTF 처리를 행하고, 그 결과 얻어진 범용 HRTF 출력 신호를 범용 고역 정보 계산부(316)에 공급한다.
- [0276] 스텝 S76에 있어서 범용 고역 정보 계산부(316)는, 범용 HRTF 처리부(315)로부터 공급된 범용 HRTF 출력 신호에 기초하여 범용 고역 정보를 계산함과 함께, 얻어진 범용 고역 정보를 압축 부호화하여, 다중화부(317)에 공급한다.
- [0277] 스텝 S77에 있어서 다중화부(317)는, 오브젝트 위치 정보 부호화부(311)로부터 공급된 부호화된 오브젝트 위치 정보, 오브젝트 신호 부호화부(313)로부터 공급된 부호화된 오브젝트 신호 및 범용 고역 정보 계산부(316)로부터 공급된 부호화된 범용 고역 정보를 다중화한다.
- [0278] 다중화부(317)는, 다중화에 의해 얻어진 출력 비트 스트림을 출력하고, 부호화 처리는 종료한다.
- [0279] 이상과 같이 하여 인코더(301)는, 범용 고역 정보를 계산하여, 출력 비트 스트림에 저장한다.
- [0280] 이렇게 함으로써, 출력 비트 스트림의 복호측에 있어서는, 범용 고역 정보를 사용하여 개인용 고역 정보를 생성할 수 있다. 이에 의해, 복호측에서는, 저비용의 장치에서도, 고품질의 오디오 재생을 행할 수 있게 된다.
- [0281] 또한, 이상에 있어서는 오디오 오브젝트의 오브젝트 신호로부터, 대역 확장의 대상이 되는 HRTF 출력 신호가 생

성되는 예에 대하여 설명했다.

- [0282] 그러나, 이에 한정되지는 않고, 예를 들어 채널 베이스의 각 채널의 오디오 신호(이하, 채널 신호라고도 칭함)로부터 HRTF 출력 신호를 생성하여, 그 HRTF 출력 신호를 대역 확장하도록 해도 된다.
- [0283] 그러한 경우, 신호 처리 장치(101)에는 렌더링 처리부(12)는 마련되지 않고, 입력 비트 스트림에는, 부호화된 채널 신호가 포함되어 있다.
- [0284] 그리고, 입력 비트 스트림에 대하여 디코드 처리부(11)가 비다중화 및 디코드 처리를 행함으로써 얻어진 멀티 채널 구성의 각 채널의 채널 신호가 버추얼라이즈 처리부(13)에 공급된다. 이들 각 채널의 채널 신호는, 각 가상 스피커의 가상 스피커 신호에 대응한다.
- [0285] 버추얼라이즈 처리부(13)는, 디코드 처리부(11)로부터 공급된 채널 신호와, HRTF 계수 기록부(122)로부터 공급된 채널마다의 개인용 HRTF 계수를 컨볼루션하고, 그 결과 얻어진 신호를 가산하는 처리를 HRTF 처리로서 행한다. 버추얼라이즈 처리부(13)는, 이러한 HRTF 처리에 의해 얻어진 HRTF 출력 신호를 대역 확장부(41)로 공급한다.
- [0286] 또한, 신호 처리 장치(101)에서 채널 신호로부터 HRTF 출력 신호가 생성되는 경우에는, 학습 장치(201)에는, 렌더링 처리부(211)가 마련되지 않고, 샘플링 주파수가 높은, 즉 고역 정보가 포함되는 채널 신호가 개인용 HRTF 처리부(212) 및 범용 HRTF 처리부(214)에 공급된다.
- [0287] 그 밖에, 예를 들어 렌더링 처리부(12)에서, HOA(High Order Ambisonics)의 렌더링 처리가 행해지도록 해도 된다.
- [0288] 그러한 경우, 예를 들어 렌더링 처리부(12)는, 디코드 처리부(11)로부터 공급된 앰비소닉스 형식, 즉 구면 조화 영역의 오디오 신호에 기초하여 렌더링 처리를 행하고, 구면 조화 영역의 가상 스피커 신호를 생성하여, 버추얼라이즈 처리부(13)에 공급한다.
- [0289] 버추얼라이즈 처리부(13)는, 렌더링 처리부(12)로부터 공급된 구면 조화 영역의 가상 스피커 신호와, HRTF 계수 기록부(122)로부터 공급된 구면 조화 영역의 개인용 HRTF 계수에 기초하여 구면 조화 영역에서 HRTF 처리를 행하고, 그 결과 얻어진 HRTF 출력 신호를 대역 확장부(41)에 공급한다. 이때, 구면 조화 영역의 HRTF 출력 신호가 대역 확장부(41)에 공급되도록 해도 되고, 필요에 따라 변환 등을 행함으로써 얻어지는 시간 영역의 HRTF 출력 신호가 대역 확장부(41)에 공급되도록 해도 된다.
- [0290] 이상과 같이, 본 기술에 의하면, 복호측(재생측)에 있어서 오브젝트 신호의 고역 정보가 아니라, 개인용 HRTF 처리 후의 신호에 대한 개인용 고역 정보를 사용하여 대역 확장 처리를 행할 수 있다.
- [0291] 또한, 이 경우, 입력 비트 스트림에 개인용 고역 정보를 다중화할 필요가 없으므로, 서버 등, 즉 인코더(301)의 스토리지 소비량을 억제할 수 있고, 또한 인코더(301)에서의 부호화 처리(인코드 처리)의 처리 시간의 증가도 억제할 수 있다.
- [0292] 또한, 재생 장치측, 즉 신호 처리 장치(101)측에서는, 디코드 처리나 렌더링 처리, 버추얼라이즈 처리를 낮은 샘플링 주파수에서 행하여, 연산량을 대폭으로 삭감할 수 있다. 이에 의해, 예를 들어 저비용의 프로세서를 채용하거나, 프로세서의 전력 사용량을 저감시키거나 할 수 있고, 스마트폰 등의 휴대 기기에서, 더 장시간, 하이레조 음원의 연속 재생을 행하는 것이 가능해진다.
- [0293] <컴퓨터의 구성예>
- [0294] 그런데, 상술한 일련의 처리는, 하드웨어에 의해 실행할 수도 있고, 소프트웨어에 의해 실행할 수도 있다. 일련의 처리를 소프트웨어에 의해 실행하는 경우에는, 그 소프트웨어를 구성하는 프로그램이 컴퓨터에 인스톨된다. 여기서, 컴퓨터에는, 전용의 하드웨어에 내장되어 있는 컴퓨터나, 각종 프로그램을 인스톨함으로써, 각종 기능을 실행하는 것이 가능한, 예를 들어 범용의 퍼스널 컴퓨터 등이 포함된다.
- [0295] 도 15는, 상술한 일련의 처리를 프로그램에 의해 실행하는 컴퓨터의 하드웨어의 구성예를 나타내는 블록도이다.
- [0296] 컴퓨터에 있어서, CPU(Central Processing Unit)(501), ROM(Read Only Memory)(502), RAM(Random Access Memory)(503)은, 버스(504)에 의해 서로 접속되어 있다.
- [0297] 버스(504)에는, 또한, 입출력 인터페이스(505)가 접속되어 있다. 입출력 인터페이스(505)에는, 입력부(506), 출력부(507), 기록부(508), 통신부(509) 및 드라이브(510)가 접속되어 있다.

- [0298] 입력부(506)는, 키보드, 마우스, 마이크로폰, 촬상 소자 등을 포함한다. 출력부(507)는, 디스플레이, 스피커 등을 포함한다. 기록부(508)는, 하드 디스크나 불휘발성의 메모리 등을 포함한다. 통신부(509)는, 네트워크 인터페이스 등을 포함한다. 드라이브(510)는, 자기 디스크, 광 디스크, 광자기 디스크, 또는 반도체 메모리 등의 리무버블 기록 매체(511)를 구동한다.
- [0299] 이상과 같이 구성되는 컴퓨터에서는, CPU(501)가, 예를 들어 기록부(508)에 기록되어 있는 프로그램을, 입출력 인터페이스(505) 및 버스(504)를 통해, RAM(503)에 로드하여 실행함으로써, 상술한 일련의 처리가 행해진다.
- [0300] 컴퓨터(CPU(501))가 실행하는 프로그램은, 예를 들어 패키지 미디어 등으로서의 리무버블 기록 매체(511)에 기록하여 제공할 수 있다. 또한, 프로그램은, 로컬 에어리어 네트워크, 인터넷, 디지털 위성 방송 등의, 유선 또는 무선의 전송 매체를 통해 제공할 수 있다.
- [0301] 컴퓨터에서는, 프로그램은, 리무버블 기록 매체(511)를 드라이브(510)에 장착함으로써, 입출력 인터페이스(505)를 통해, 기록부(508)에 인스톨할 수 있다. 또한, 프로그램은, 유선 또는 무선의 전송 매체를 통해, 통신부(509)에서 수신하고, 기록부(508)에 인스톨할 수 있다. 그 밖에, 프로그램은, ROM(502)이나 기록부(508)에, 미리 인스톨해 둘 수 있다.
- [0302] 또한, 컴퓨터가 실행하는 프로그램은, 본 명세서에서 설명하는 순서에 따라 시계열로 처리가 행해지는 프로그램 이어도 되고, 병렬로, 혹은 호출이 행해졌을 때 등의 필요한 타이밍에 처리가 행해지는 프로그램이어도 된다.
- [0303] 또한, 본 기술의 실시 형태는, 상술한 실시 형태에 한정되는 것은 아니고, 본 기술의 요지를 일탈하지 않는 범위에 있어서 다양한 변경이 가능하다.
- [0304] 예를 들어, 본 기술은, 하나의 기능을 네트워크를 통해 복수의 장치에서 분담, 공동하여 처리하는 클라우드 컴퓨팅의 구성을 취할 수 있다.
- [0305] 또한, 상술한 흐름도에서 설명한 각 스텝은, 하나의 장치에서 실행하는 것 외에, 복수의 장치에서 분담하여 실행할 수 있다.
- [0306] 또한, 하나의 스텝에 복수의 처리가 포함되는 경우에는, 그 하나의 스텝에 포함되는 복수의 처리는, 하나의 장치에서 실행하는 것 외에, 복수의 장치에서 분담하여 실행할 수 있다.
- [0307] 또한, 본 기술은, 이하의 구성으로 하는 것도 가능하다.
- [0308] (1)
- [0309] 입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하는 디코딩 처리부와,
- [0310] 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는 대역 확장부
- [0311] 를 구비하는 신호 처리 장치.
- [0312] (2)
- [0313] 상기 제1 고역 정보에 기초하여 상기 제2 고역 정보를 생성하는 고역 정보 생성부를 더 구비하는
- [0314] (1)에 기재된 신호 처리 장치.
- [0315] (3)
- [0316] 상기 제1 고역 정보는, 제1 계수를 사용하여 상기 신호 처리를 행함으로써 얻어지는 상기 제2 오디오 신호의 대역 확장을 위한 고역 정보이고,
- [0317] 상기 제2 고역 정보는, 제2 계수를 사용하여 상기 신호 처리를 행함으로써 얻어지는 상기 제2 오디오 신호의 대역 확장을 위한 고역 정보이고,
- [0318] 상기 대역 확장부는, 상기 제1 오디오 신호, 상기 메타데이터 및 상기 제2 계수에 기초하여 상기 신호 처리를 행함으로써 얻어진 상기 제2 오디오 신호와, 상기 제2 고역 정보에 기초하여 상기 대역 확장 처리를 행하는
- [0319] (2)에 기재된 신호 처리 장치.

- [0320] (4)
- [0321] 상기 고역 정보 생성부는, 상기 제1 고역 정보, 상기 제1 계수 및 상기 제2 계수에 기초하여, 상기 제2 고역 정보를 생성하는
- [0322] (3)에 기재된 신호 처리 장치.
- [0323] (5)
- [0324] 상기 고역 정보 생성부는, 미리 기계 학습에 의해 생성된 계수와, 상기 제1 고역 정보, 상기 제1 계수 및 상기 제2 계수에 기초하는 연산을 행함으로써, 상기 제2 고역 정보를 생성하는
- [0325] (3) 또는 (4)에 기재된 신호 처리 장치.
- [0326] (6)
- [0327] 상기 연산은 뉴럴 네트워크에 의한 연산인
- [0328] (5)에 기재된 신호 처리 장치.
- [0329] (7)
- [0330] 상기 제1 계수는 범용의 계수이고, 상기 제2 계수는 유저마다의 계수인
- [0331] (3) 내지 (6) 중 어느 한 항에 기재된 신호 처리 장치.
- [0332] (8)
- [0333] 상기 제1 계수 및 상기 제2 계수는 HRTF 계수인
- [0334] (7)에 기재된 신호 처리 장치.
- [0335] (9)
- [0336] 상기 제1 계수를 기록하는 계수 기록부를 더 구비하는
- [0337] (3) 내지 (8) 중 어느 한 항에 기재된 신호 처리 장치.
- [0338] (10)
- [0339] 상기 신호 처리를 행하여 상기 제2 오디오 신호를 생성하는 신호 처리부를 더 구비하는
- [0340] (1) 내지 (9) 중 어느 한 항에 기재된 신호 처리 장치.
- [0341] (11)
- [0342] 상기 신호 처리는 벡터얼라이즈 처리를 포함하는 처리인
- [0343] (10)에 기재된 신호 처리 장치.
- [0344] (12)
- [0345] 상기 신호 처리는 렌더링 처리를 포함하는 처리인
- [0346] (11)에 기재된 신호 처리 장치.
- [0347] (13)
- [0348] 상기 제1 오디오 신호는, 오디오 오브젝트의 오브젝트 신호, 또는 채널 베이스의 오디오 신호인
- [0349] (1) 내지 (12) 중 어느 한 항에 기재된 신호 처리 장치.
- [0350] (14)
- [0351] 신호 처리 장치가,
- [0352] 입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하고,
- [0353] 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기

제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는

- [0354] 신호 처리 방법.
- [0355] (15)
- [0356] 입력 비트 스트림을 제1 오디오 신호와, 상기 제1 오디오 신호의 메타데이터와, 대역 확장을 위한 제1 고역 정보에 비다중화하고,
- [0357] 상기 제1 오디오 신호 및 상기 메타데이터에 기초하여 신호 처리를 행함으로써 얻어진 제2 오디오 신호와, 상기 제1 고역 정보에 기초하여 생성된 제2 고역 정보에 기초하여 대역 확장 처리를 행하여, 출력 오디오 신호를 생성하는
- [0358] 스템을 포함하는 처리를 컴퓨터에 실행시키는 프로그램.
- [0359] (16)
- [0360] 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하는 제1 고역 정보 계산부와,
- [0361] 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하는 제2 고역 정보 계산부와,
- [0362] 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는 고역 정보 학습부를 구비하는 학습 장치.
- [0363] (17)
- [0364] 상기 계수 데이터는 뉴럴 네트워크를 구성하는 계수인
- [0365] (16)에 기재된 학습 장치.
- [0366] (18)
- [0367] 상기 제1 계수는 범용의 계수이고, 상기 제2 계수는 유저마다의 계수인
- [0368] (16) 또는 (17)에 기재된 학습 장치.
- [0369] (19)
- [0370] 상기 신호 처리는 버추얼라이즈 처리를 포함하는 처리이고,
- [0371] 상기 제1 계수 및 상기 제2 계수는 HRTF 계수인
- [0372] (18)에 기재된 학습 장치.
- [0373] (20)
- [0374] 상기 신호 처리는 렌더링 처리를 포함하는 처리인
- [0375] (19)에 기재된 학습 장치.
- [0376] (21)
- [0377] 상기 제1 오디오 신호는, 오디오 오브젝트의 오브젝트 신호, 또는 채널 베이스의 오디오 신호인
- [0378] (16) 내지 (19) 중 어느 한 항에 기재된 학습 장치.
- [0379] (22)
- [0380] 학습 장치가,
- [0381] 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위

한 제1 고역 정보를 생성하고,

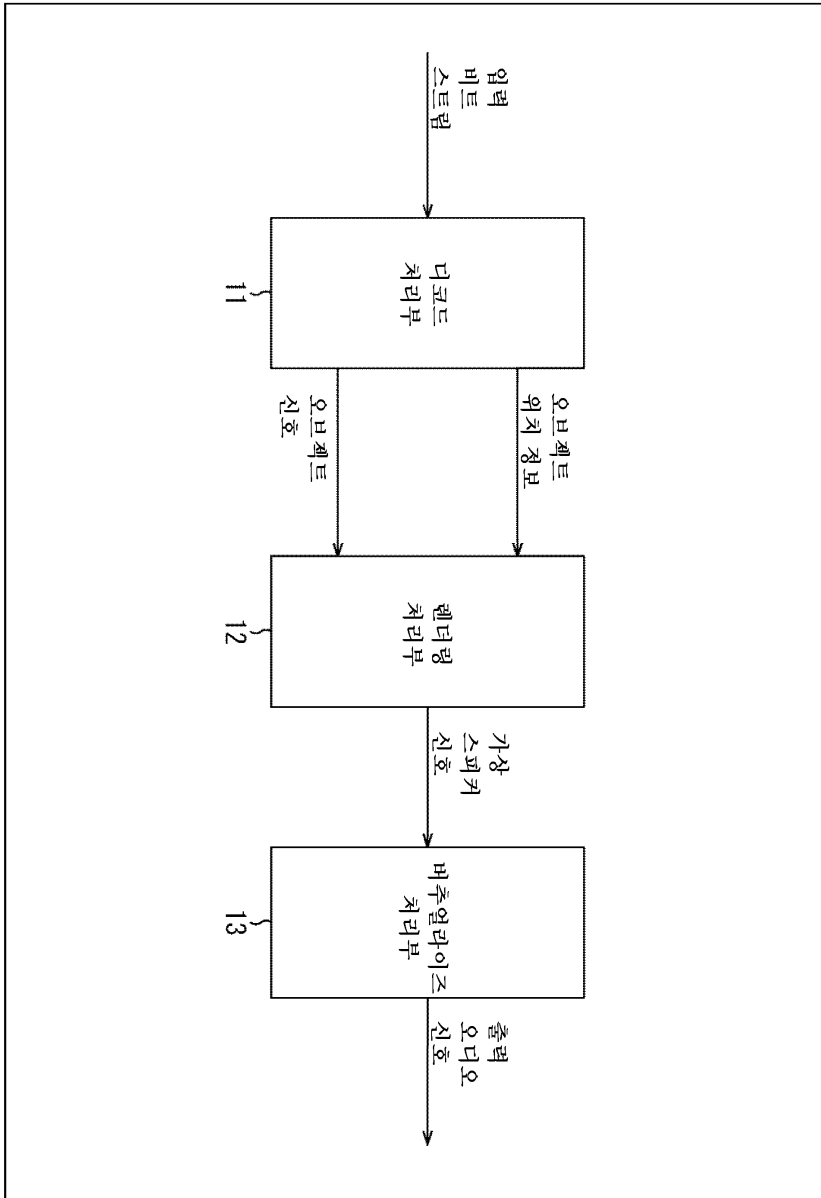
- [0383] 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하고,
- [0384] 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는
- [0385] 학습 방법.
- [0386] (23)
- [0387] 제1 오디오 신호와 제1 계수에 기초하는 신호 처리에 의해 생성된 제2 오디오 신호에 기초하여, 대역 확장을 위한 제1 고역 정보를 생성하고,
- [0388] 상기 제1 오디오 신호와 제2 계수에 기초하는 상기 신호 처리에 의해 생성된 제3 오디오 신호에 기초하여, 대역 확장을 위한 제2 고역 정보를 생성하고,
- [0389] 상기 제1 계수, 상기 제2 계수, 상기 제1 고역 정보 및 상기 제2 고역 정보에 기초하여, 상기 제2 고역 정보를 교사 데이터로 하는 학습을 행하여, 상기 제1 계수, 상기 제2 계수 및 상기 제1 고역 정보로부터 상기 제2 고역 정보를 얻기 위한 계수 데이터를 생성하는
- [0390] 스템을 포함하는 처리를 컴퓨터에 실행시키는 프로그램.

**부호의 설명**

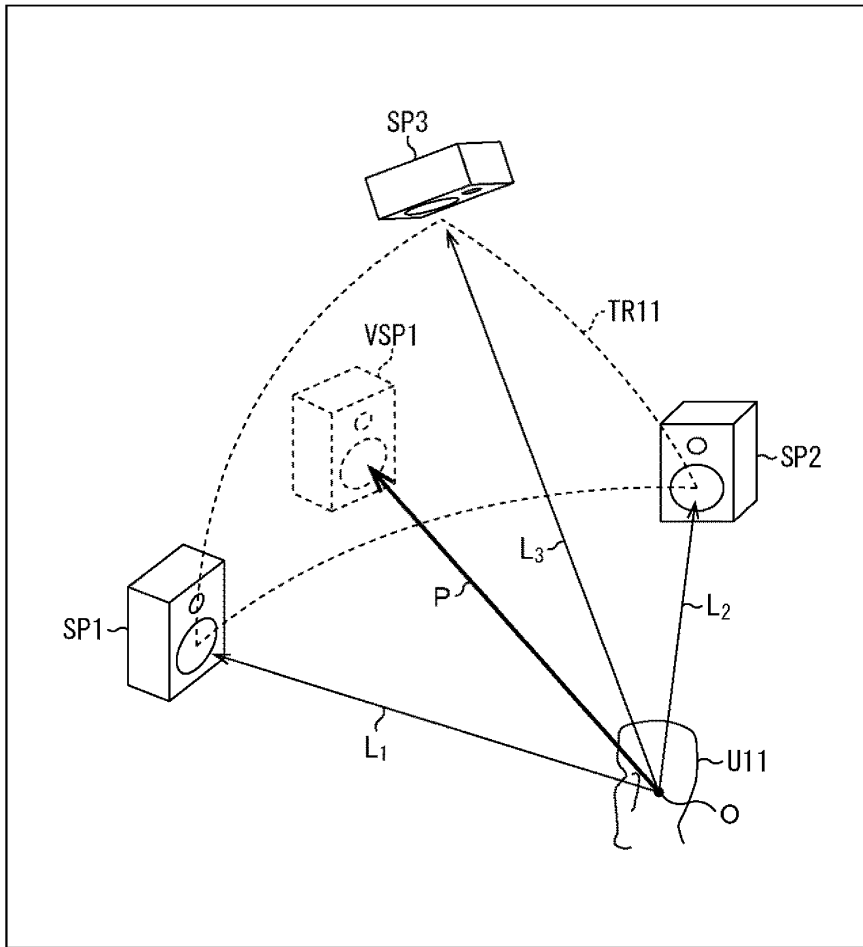
- [0391] 11: 디코드 처리부
- 12: 렌더링 처리부
- 13: 버추얼라이즈 처리부
- 41: 대역 확장부
- 101: 신호 처리 장치
- 121: 개인용 고역 정보 생성부

도면

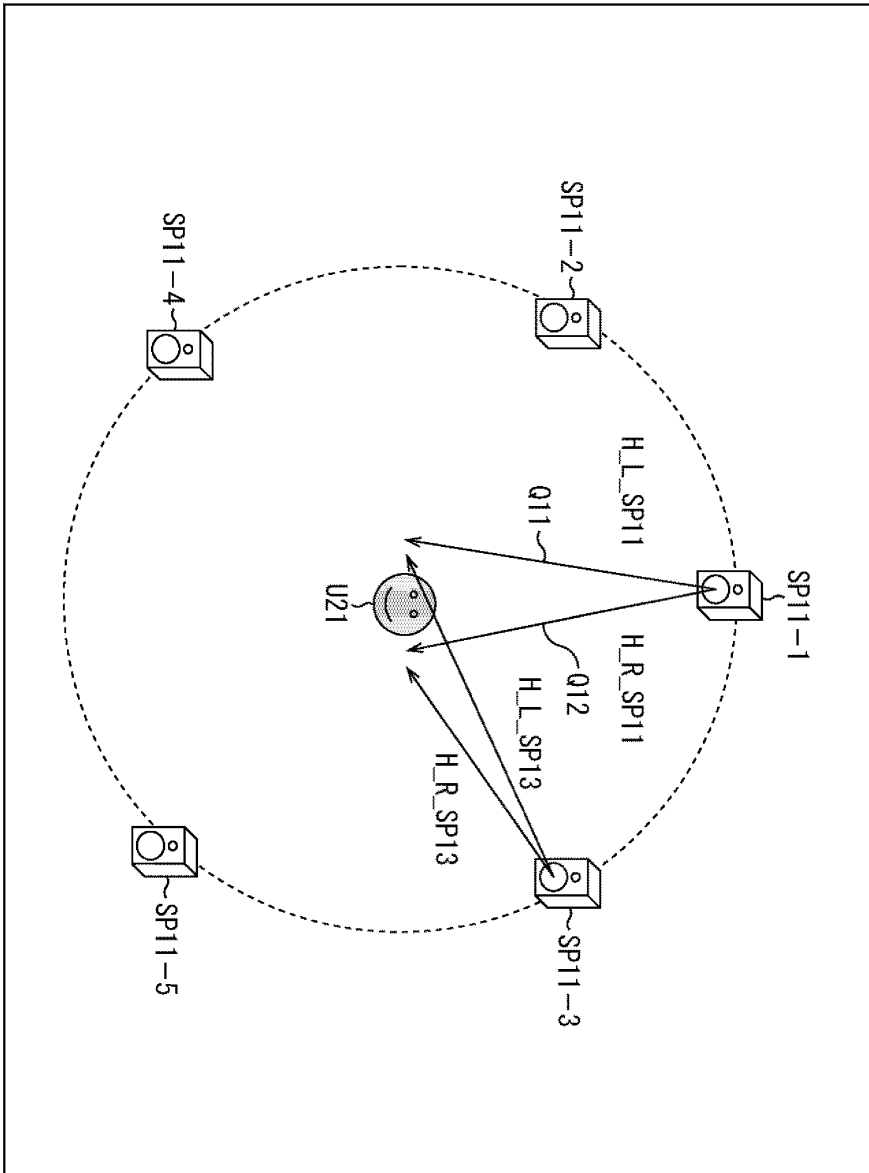
도면1



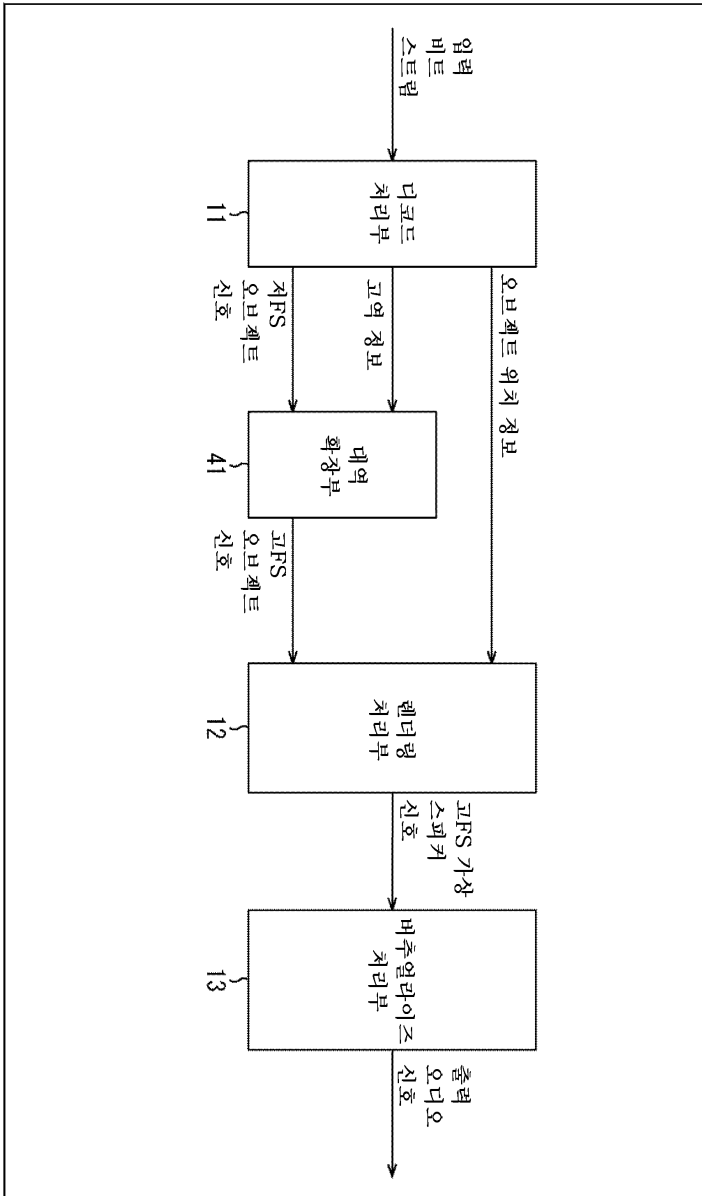
도면2



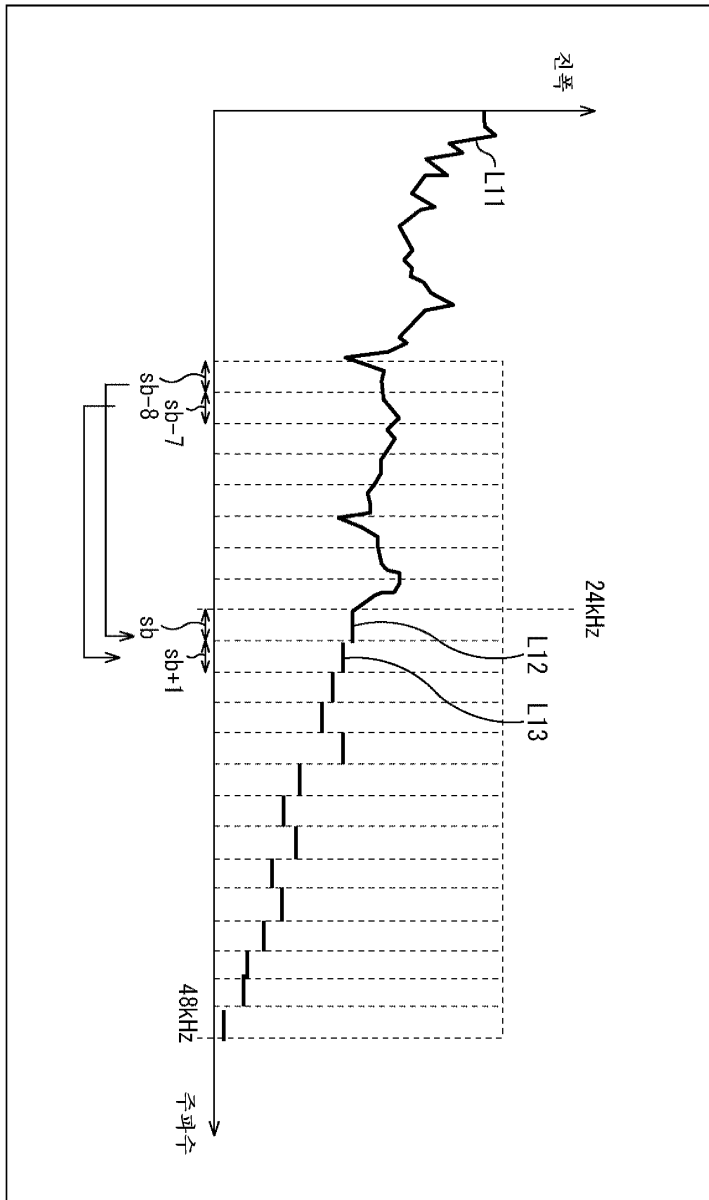
도면3



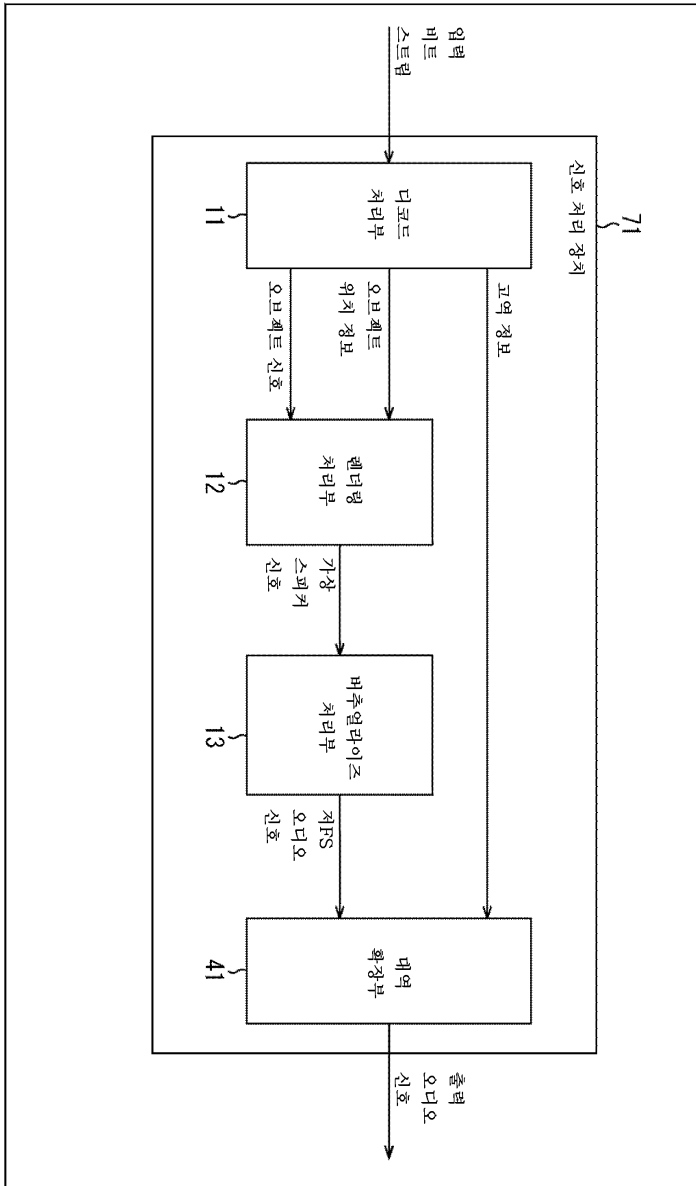
도면4



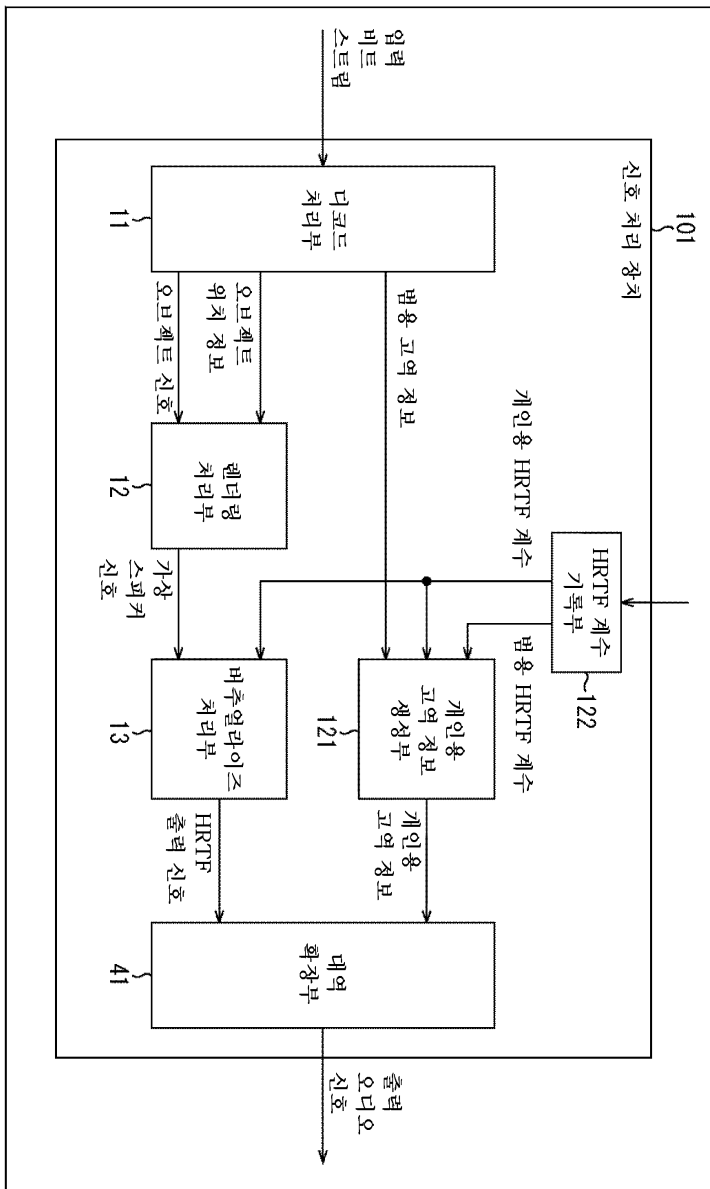
도면5



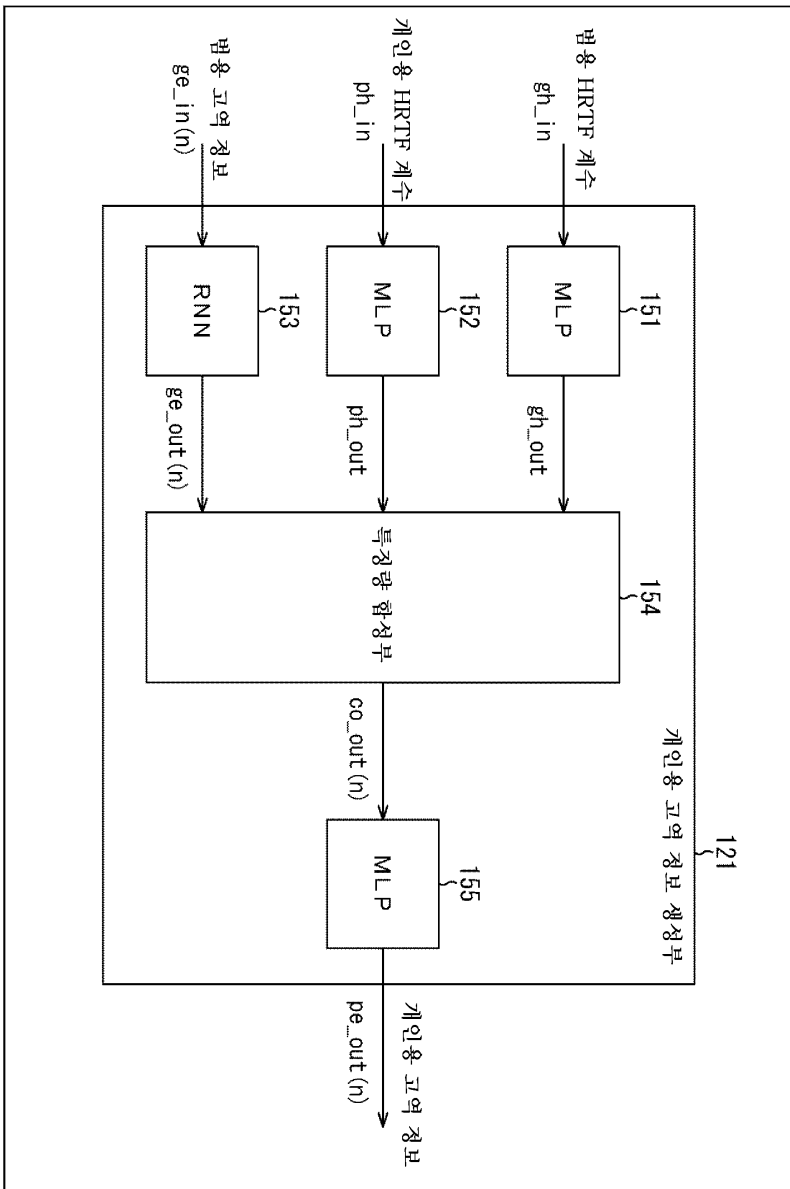
도면6



도면7



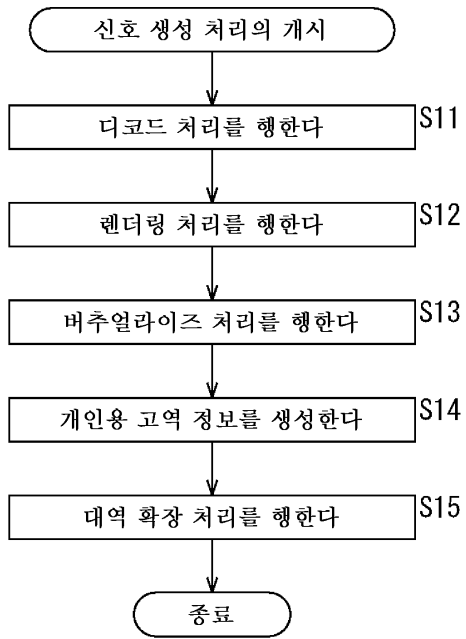
도면8



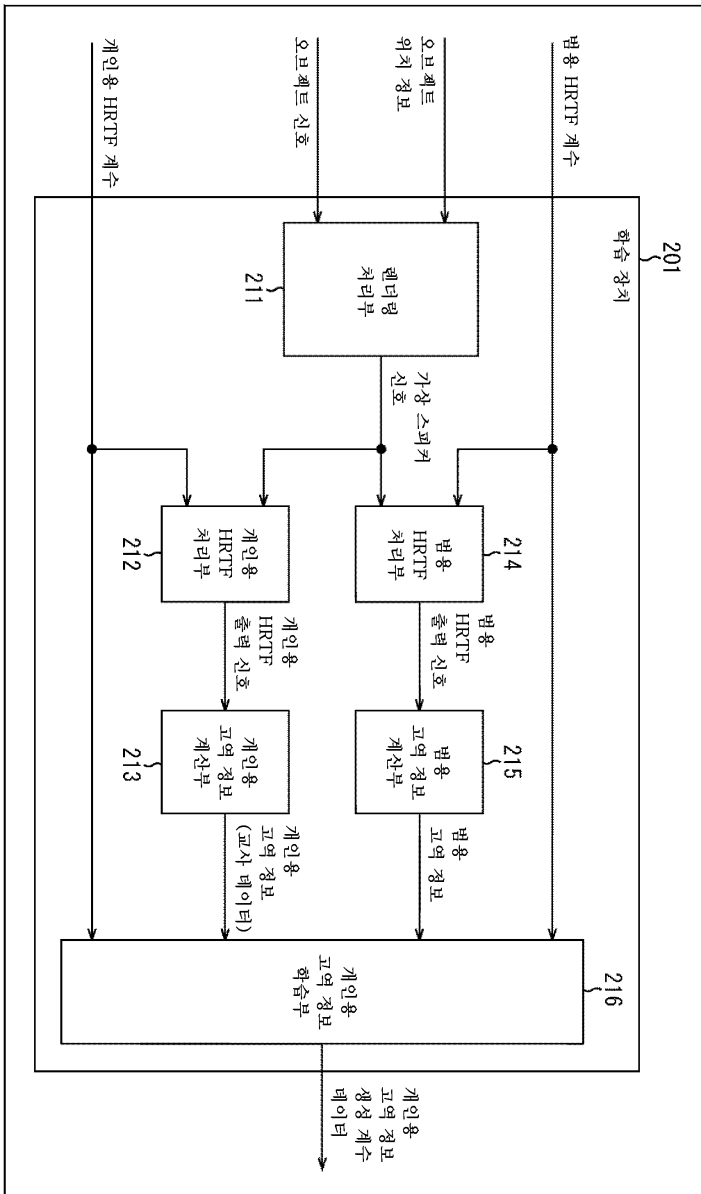
도면9

Syntax	No. of bits	Mnemonic
<pre> for ( i=0; i &lt; num_objects; i++) {   object_compressed_data(i) } for ( i=0; i &lt; num_objects; i++) {   position_azimuth[i];   position_elevation[i];   position_radius[i]; } for(i = 0; i &lt; num_output; i++){   output_bwe_data(i) }                 </pre>	<p>8 6 4</p>	<p>tcimsbf tcimsbf uimsbf</p>

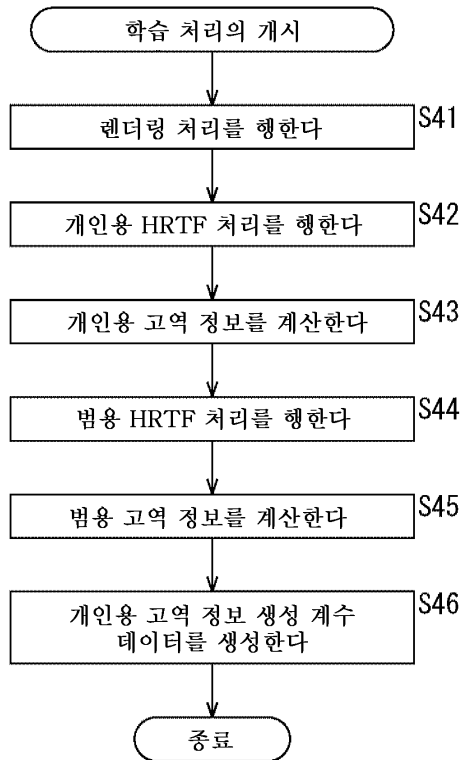
도면10



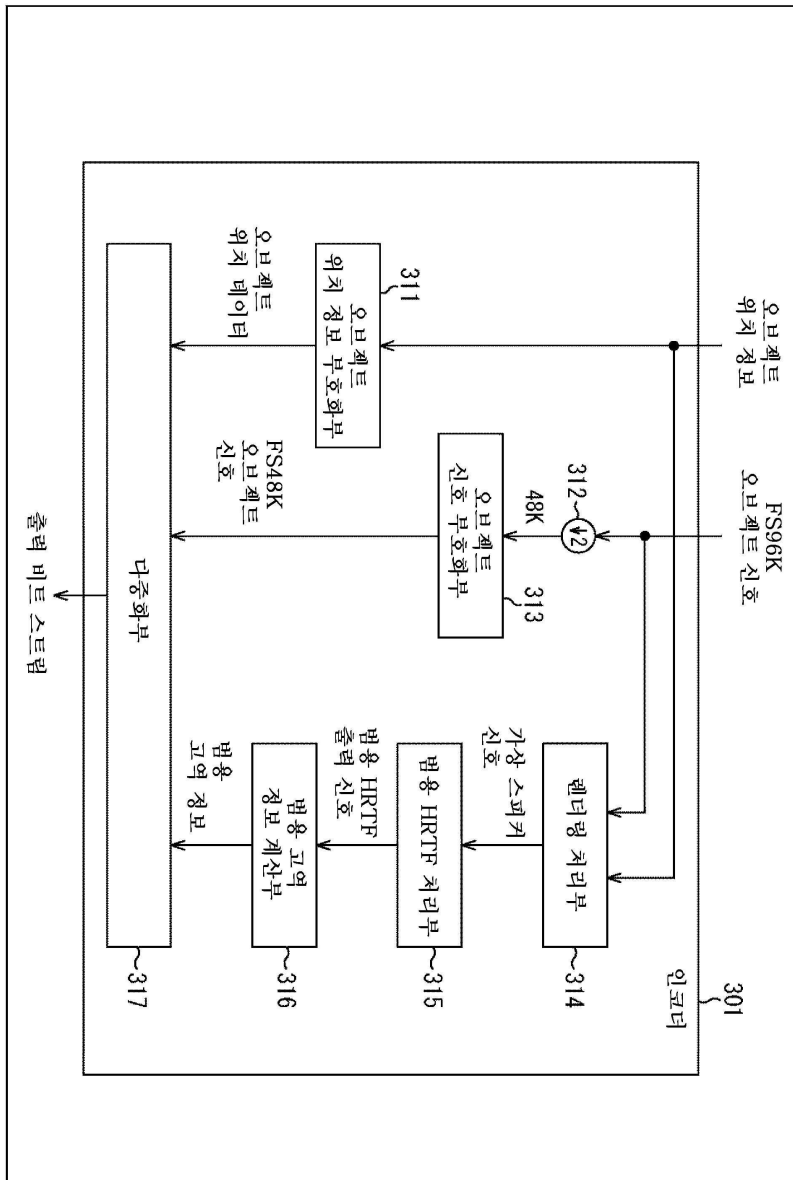
도면11



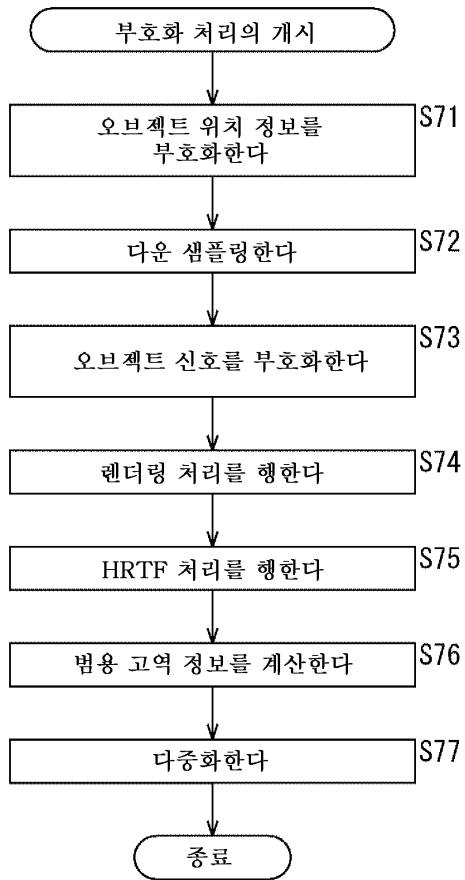
도면12



도면13



도면14



도면15

