



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0161342  
(43) 공개일자 2024년11월12일

(51) 국제특허분류(Int. Cl.)  
G10L 19/038 (2013.01) G10L 19/00 (2006.01)  
G10L 19/06 (2006.01) G10L 25/30 (2013.01)  
(52) CPC특허분류  
G10L 19/038 (2013.01)  
G10L 19/0017 (2013.01)  
(21) 출원번호 10-2023-0058335  
(22) 출원일자 2023년05월04일  
심사청구일자 2024년05월08일

(71) 출원인  
한국전자통신연구원  
대전광역시 유성구 가정로 218 (가정동)  
연세대학교 산학협력단  
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)  
(72) 발명자  
장인선  
대전광역시 유성구 가정로 218 (가정동)  
백승권  
대전광역시 유성구 가정로 218 (가정동)  
(74) 대리인  
특허법인 무한

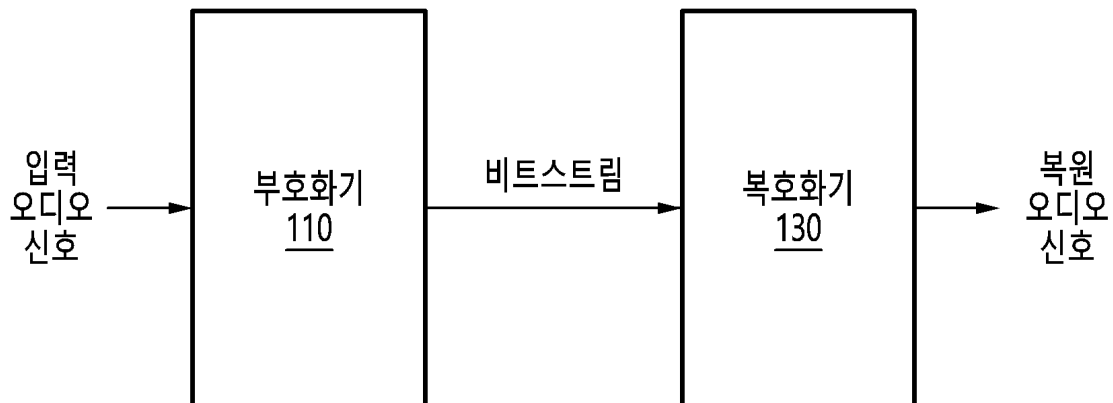
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 오디오 신호의 부호화/복호화 방법 및 장치

(57) 요약

오디오 신호의 부호화/복호화 방법 및 장치가 개시된다. 일 실시예에 따른, 부호화 방법은, 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하는 동작과, 상기 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화하는 동작과, 상기 주파수 영역의 오디오 신호를 상기 양자화된 에너지에 따라 정규화한 정규 신호(뒷면에 계속)

대표도 - 도1



호를 생성하는 동작과, 상기 정규 신호 및 상기 입력 오디오 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터(feature vector)를 획득하는 동작과, 상기 특징 벡터를 양자화하는 동작과,

양자화된 특징 벡터에 기초하여, 상기 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득하는 동작과, 상기 스케일 값에 기초하여 상기 정규 신호를 스케일링한 조정 신호를 양자화하는 동작과, 상기 양자화된 에너지, 상기 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bit stream)들을 출력하는 동작을 포함할 수 있다.

(52) CPC특허분류

**G10L 19/06** (2013.01)

**G10L 25/30** (2013.01)

(72) 발명자

**성종모**

대전광역시 유성구 가정로 218 (가정동)

**이태진**

대전광역시 유성구 가정로 218 (가정동)

**임우택**

대전광역시 유성구 가정로 218 (가정동)

**조병호**

대전광역시 유성구 가정로 218 (가정동)

**강홍구**

서울특별시 서대문구 연세로 50, 연세대학교 제2공학관 601호 (신촌동)

**김병현**

서울특별시 강남구 선릉로126길 22, 111동 902호 (삼성동, 롯데캐슬프리미어아파트)

**이지현**

서울특별시 서초구 효령로5길 9, 503호 (방배동, 방배리치빌)

**임형섭**

서울특별시 성북구 성북로8가길 14, 201호 (성북동)

이 발명을 지원한 국가연구개발사업

과제고유번호

1711173562

과제번호

22ZH1200

부처명

과학기술정보통신부

과제관리(전문)기관명

한국전자통신연구원

연구사업명

한국전자통신연구원연구운영비지원(주요사업비)

연구과제명

초실감 입체공간 미디어·콘텐츠 원천기술연구

기 여 율

1/1

과제수행기관명

한국전자통신연구원

연구기간

2022.01.01 ~ 2022.12.31

## 명세서

### 청구범위

#### 청구항 1

시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하는 동작;  
 상기 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화하는 동작;  
 상기 주파수 영역의 오디오 신호를 상기 양자화된 에너지에 따라 정규화한 정규 신호를 생성하는 동작;  
 상기 정규 신호 및 상기 입력 오디오 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터(feature vector)를 획득하는 동작;  
 상기 특징 벡터를 양자화하는 동작;  
 양자화된 특징 벡터에 기초하여, 상기 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득하는 동작;  
 상기 스케일 값에 기초하여 상기 정규 신호를 스케일링한 조정 신호를 양자화하는 동작; 및  
 상기 양자화된 에너지, 상기 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bit stream)들을 출력하는 동작;  
 을 포함하는, 부호화 방법.

#### 청구항 2

제1항에 있어서,  
 상기 특징 벡터를 획득하는 동작은,  
 상기 입력 오디오 신호에 기초하여, 주파수 영역내 상기 입력 오디오 신호의 크기 스펙트럼(magnitude spectrum)을 획득하는 동작; 및  
 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 특징 벡터를 획득하는 동작  
 을 포함하는, 부호화 방법.

#### 청구항 3

제2항에 있어서,  
 상기 특징 벡터를 획득하는 동작은,  
 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 생성하는 동작; 및  
 상기 잠재 표현에 기초하여, 상기 특징 벡터를 계산하는 동작  
 을 포함하는, 부호화 방법.

#### 청구항 4

제1항에 있어서,  
 상기 스케일 값을 획득하는 동작은,

상기 양자화된 특징 벡터에 기초하여, 상기 주파수 밴드 별로 스케일 값을 획득하는 동작을 포함하고,  
상기 스케일 값의 차원 수는,  
상기 주파수 밴드의 총 밴드 수와 일치하는 것이고,  
상기 양자화된 특징 벡터의 차원 수는,  
상기 특징 벡터의 차원 수와 일치하는 것인,  
부호화 방법.

## 청구항 5

제1항에 있어서,  
상기 조정 신호를 양자화하는 동작은,  
상기 정규 신호를 상기 스케일 값에 따라 스케일링하여, 상기 조정 신호를 생성하는 동작을 포함하는, 부호화 방법.

## 청구항 6

제1항에 있어서,  
상기 비트 스트림들을 출력하는 동작은,  
상기 양자화된 특징 벡터를 부호화하여 제1 비트 스트림을 출력하는 동작;  
상기 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력하는 동작; 및  
상기 양자화된 에너지를 부호화하여, 제3 비트 스트림을 출력하는 동작을 포함하는, 부호화 방법

## 청구항 7

부호화기로부터 비트 스트림들을 수신하는 동작;  
양자화된 특징 벡터를 부호화한 제1 비트 스트림에 기초하여, 복원 조정 신호를 역 스케일링 하는데 사용되는 스케일 값을 획득하는 동작;  
상기 스케일 값, 및 양자화된 조정 신호를 부호화한 제2 비트 스트림에 기초하여, 복원 정규 신호를 생성하는 동작;  
양자화된 에너지를 부호화한 제3 비트 스트림, 및 상기 복원 정규 신호에 기초하여, 주파수 영역의 복원 오디오 신호를 획득하는 동작; 및  
상기 주파수 영역의 복원 오디오 신호에 기초하여, 시간 영역의 복원 오디오 신호를 출력하는 동작을 포함하는, 복호화 방법.

## 청구항 8

제7항에 있어서,  
상기 스케일 값을 획득하는 동작은,

상기 제1 비트 스트림을 복호화 하여 양자화된 특징 벡터를 획득하는 동작; 및  
상기 양자화된 특징 벡터로부터 상기 스케일 값을 계산하는 동작  
을 포함하는, 복호화 방법.

#### 청구항 9

제7항에 있어서,  
상기 복원 정규 신호를 생성하는 동작은,  
상기 제2 비트 스트림을 복호화 하여, 복원 조정 신호를 생성하는 동작; 및  
상기 복원 조정 신호를 상기 스케일 값에 따라 역 스케일링(inverse scaling)하는 동작;  
을 포함하는, 복호화 방법.

#### 청구항 10

제7항에 있어서,  
상기 주파수 영역의 복원 오디오 신호를 획득하는 동작은,  
상기 제3 비트 스트림에 기초하여, 상기 주파수 영역의 복원 오디오 신호의 주파수 밴드의 복원 에너지를 출력  
하는 동작; 및  
상기 복원 정규 신호를 상기 복원 에너지에 따라 역 정규화(denormalization)하는 동작  
을 포함하는, 복호화 방법.

#### 청구항 11

하나 이상의 인스트럭션을 저장하는 메모리; 및  
상기 인스트럭션을 실행시키기 위한 프로세서  
를 포함하고,  
상기 인스트럭션이 실행될 때, 상기 프로세서는 복수의 동작들을 수행하고,  
상기 복수의 동작들은,  
시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하는 동작;  
상기 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화하는 동작;  
상기 주파수 영역의 오디오 신호를 상기 양자화된 에너지에 따라 정규화한 정규 신호를 생성하는 동작;  
상기 정규 신호 및 상기 입력 오디오 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 포함하는 특  
징 벡터(feature vector)를 획득하는 동작;  
상기 특징 벡터를 양자화하는 동작;  
양자화된 특징 벡터에 기초하여, 상기 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득  
하는 동작;  
상기 스케일 값에 기초하여 상기 정규 신호를 스케일링한 조정 신호를 양자화하는 동작; 및  
상기 양자화된 에너지, 상기 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bit stream)들을 출력하는 동작;

을 포함하는, 부호화 장치.

## 청구항 12

제11항에 있어서,

상기 특징 벡터를 획득하는 동작은,

상기 입력 오디오 신호에 기초하여, 주파수 영역내 상기 입력 오디오 신호의 크기 스펙트럼(magnitude spectrum)을 획득하는 동작; 및

상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 특징 벡터를 획득하는 동작

을 포함하는, 부호화 장치.

## 청구항 13

제12항에 있어서,

상기 특징 벡터를 획득하는 동작은,

상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 생성하는 동작; 및

상기 잠재 표현에 기초하여, 상기 특징 벡터를 계산하는 동작

을 포함하는, 부호화 장치.

## 청구항 14

제11항에 있어서,

상기 스케일 값을 획득하는 동작은,

상기 양자화된 특징 벡터에 기초하여, 상기 주파수 밴드 별로 스케일 값을 획득하는 동작

을 포함하고,

상기 스케일 값의 차원 수는,

상기 주파수 밴드의 총 밴드 수와 일치하는 것이고,

상기 양자화된 특징 벡터의 차원 수는,

상기 특징 벡터의 차원 수와 일치하는 것인,

부호화 장치.

## 청구항 15

제11항에 있어서,

상기 조정 신호를 양자화하는 동작은,

상기 정규 신호를 상기 스케일 값에 따라 스케일링하여, 상기 조정 신호를 생성하는 동작

을 포함하는, 부호화 장치.

**청구항 16**

제11항에 있어서,  
 상기 비트 스트림들을 출력하는 동작은,  
 상기 양자화된 특징 벡터를 부호화하여 제1 비트 스트림을 출력하는 동작;  
 상기 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력하는 동작; 및  
 상기 양자화된 에너지를 부호화하여, 제3 비트 스트림을 출력하는 동작  
 을 포함하는, 부호화 장치.

**발명의 설명****기술 분야**

[0001] 아래 개시는 오디오 신호의 부호화/복호화 방법 및 장치에 관한 것이다.

**배경 기술**

[0003] 오디오 신호를 압축하고 복원하는 기술은 오디오 신호를 효율적으로 장치에 저장하거나 통신망을 통해 빠르게 신호를 송수신하기 위해 광범위하게 사용되고 있다. 특히, 인터넷을 통해 오고 가는 미디어 신호의 양이 점차 늘어남에 따라 보다 적은 양의 압축된 데이터만으로 복원 시의 왜곡을 최소화할 수 있는 오디오 부호화(encoding) 및 복호화(decoding) 기술에 대한 관심이 더욱 커지고 있다.

[0004] 기존에 표준화된 오디오 부호화 기술은 심리 음향학적(psychoacoustics) 지식을 이용하여 복호화로부터 발생하는 신호 왜곡이 일반적인 청취 능력을 가진 사람으로 하여금 가능한 인지되지 못하도록 설계되었다. 특히, 특정한 고유 주파수를 가지는 성분의 세기가 큰 경우 인접한 고유 주파수의 신호 성분이 잘 들리지 않게 되는 마스킹 효과(masking effect)를 활용하여 양자화(quantization) 왜곡이 실제 청취 시에는 인지되기 어렵도록 각 주파수 성분 별로 마스킹 효과의 정도를 미리 계산한 뒤 이에 따라 각 주파수 성분의 양자화 단계를 조절한다. 이를 가장 효과적으로 활용하기 위해 부호화기의 입력으로 들어오는 시간 영역(time-domain) 신호는 변형 이산 코사인 변환(modified discrete cosine transform; MDCT) 혹은 이산 푸리에 변환(discrete Fourier transform; DFT) 의해 주파수 영역(frequency-domain) 신호로 변환되며, 마스킹 효과에 대한 분석이 이루어진 뒤 실제 양자화 과정을 거쳐 비트열(bitstream)로 변환되고, 복호화기에서 역변환을 거쳐 시간 영역 신호로 복원된다.

[0005] 이와 같은 방식으로 구현된 장치들은 이미 상용화되어 다양한 영역에서 활용되고 있으나, 각 장치를 구성하는 요소들은 독립적으로 최적화될 수밖에 없어 해당 구성 요소들 간의 입출력을 조절함에 있어 설계상 어려움이 따른다. 이러한 한계점을 극복하기 위해 부호화기, 양자화기, 그리고 복호화기의 모든 장치 구성 요소를 미분 가능한 심층신경망 구조로 대체하고, 최종 복호화 이후의 왜곡을 최소화하도록 전체 시스템을 동시에 최적화하는 종단 간(end-to-end) 학습 모델에 대한 연구가 최근 다양하게 이루어지고 있다. 이와 같은 접근 방식은 해석이 용이한 선형(linear) 시스템에 장치 구성을 의존하는 대신 복잡한 비선형(nonlinear) 시스템의 원활한 도입을 가능케 한다는 점에서 기존의 규칙 기반(rule-based) 부호화기 및 복호화기를 향상시키는 데 기여할 수 있을 것으로 기대받고 있다.

[0006] 위에서 설명한 배경기술은 발명자가 본원의 개시 내용을 도출하는 과정에서 보유하거나 습득한 것으로서, 반드시 본 출원 전에 일반 공중에 공개된 공지기술이라고 할 수는 없다.

**발명의 내용****해결하려는 과제**

[0008] 특정한 주파수 성분의 세기를 선택적으로 조절하기 위해서, 주파수 영역에서 신호를 압축하기 위한 심층 신경망

기본 종단간 장치 구성 방식을 구현하는 기술이 요구된다.

- [0009] 일 실시예는 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하여, 주파수 영역의 오디오 신호의 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터를 획득하는 기술을 제공할 수 있다.
- [0010] 일 실시예는, 특징벡터를 양자화하여, 양자화된 특징벡터로부터 정규 신호를 스케일링 하는데 사용되는 스케일 값을 획득하는 기술을 제공할 수 있다.
- [0011] 다만, 기술적 과제는 상술한 기술적 과제들로 한정되는 것은 아니며, 또 다른 기술적 과제들이 존재할 수 있다.

## 과제의 해결 수단

- [0013] 일 실시예에 따른 오디오 신호의 부호화 방법은 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하는 동작과, 상기 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화하는 동작과, 상기 주파수 영역의 오디오 신호를 상기 양자화된 에너지에 따라 정규화한 정규 신호를 생성하는 동작과, 상기 정규 신호 및 상기 입력 오디오 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터(feature vector)를 획득하는 동작과, 상기 특징 벡터를 양자화하는 동작과, 양자화된 특징 벡터에 기초하여, 상기 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득하는 동작과, 상기 스케일 값에 기초하여 상기 정규 신호를 스케일링한 조정 신호를 양자화하는 동작과, 상기 양자화된 에너지, 상기 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bit stream)들을 출력하는 동작을 포함할 수 있다.
- [0014] 상기 특징 벡터를 획득하는 동작은 상기 입력 오디오 신호에 기초하여, 주파수 영역내 상기 입력 오디오 신호의 크기 스펙트럼(magnitude spectrum)을 획득하는 동작과, 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 특징 벡터를 획득하는 동작을 포함할 수 있다.
- [0015] 상기 특징 벡터를 획득하는 동작은 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 생성하는 동작과, 상기 잠재 표현에 기초하여, 상기 특징 벡터를 계산하는 동작을 포함할 수 있다.
- [0016] 상기 스케일 값을 획득하는 동작은 상기 양자화된 특징 벡터에 기초하여, 상기 주파수 밴드 별로 스케일 값을 획득하는 동작을 포함할 수 있다. 상기 스케일 값의 차원 수는, 상기 주파수 밴드의 총 밴드 수와 일치하는 것일 수 있고, 상기 양자화된 특징 벡터의 차원 수는, 상기 특징 벡터의 차원 수와 일치하는 것일 수 있다.
- [0017] 상기 조정 신호를 양자화하는 동작은 상기 정규 신호를 상기 스케일 값에 따라 스케일링하여, 상기 조정 신호를 생성하는 동작을 포함할 수 있다.
- [0018] 상기 비트 스트림들을 출력하는 동작은 상기 양자화된 특징 벡터를 부호화하여 제1 비트 스트림을 출력하는 동작과, 상기 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력하는 동작과, 상기 양자화된 에너지를 부호화하여, 제3 비트 스트림을 출력하는 동작을 포함할 수 있다.
- [0019] 일 실시예에 따른 오디오 신호의 복호화 방법은 부호화기로부터 비트 스트림들을 수신하는 동작과, 양자화된 특징 벡터를 부호화한 제1 비트 스트림에 기초하여, 복원 조정 신호를 역 스케일링 하는데 사용되는 스케일 값을 획득하는 동작과, 상기 스케일 값, 및 양자화된 조정 신호를 부호화한 제2 비트 스트림에 기초하여, 복원 정규 신호를 생성하는 동작과, 양자화된 에너지를 부호화한 제3 비트 스트림, 및 상기 복원 정규 신호에 기초하여, 주파수 영역의 복원 오디오 신호를 획득하는 동작과, 상기 주파수 영역의 복원 오디오 신호에 기초하여, 시간 영역의 복원 오디오 신호를 출력하는 동작을 포함할 수 있다.
- [0020] 상기 스케일 값을 획득하는 동작은 상기 제1 비트 스트림을 복호화 하여 양자화된 특징 벡터를 획득하는 동작과, 상기 양자화된 특징 벡터로부터 상기 스케일 값을 계산하는 동작을 포함할 수 있다.
- [0021] 상기 복원 정규 신호를 생성하는 동작은 상기 제2 비트 스트림을 복호화 하여, 복원 조정 신호를 생성하는 동작과, 상기 복원 조정 신호를 상기 스케일 값에 따라 역 스케일링(inverse scaling)하는 동작을 포함할 수 있다.
- [0022] 상기 주파수 영역의 복원 오디오 신호를 획득하는 동작은 상기 제3 비트 스트림에 기초하여, 상기 주파수 영역의 복원 오디오 신호의 주파수 밴드의 복원 에너지를 출력하는 동작과, 상기 복원 정규 신호를 상기 복원 에너지에 따라 역 정규화(denormalization)하는 동작을 포함할 수 있다.
- [0023] 일 실시예에 따른 부호화 장치는 하나 이상의 인스트럭션을 저장하는 메모리와 상기 인스트럭션을 실행시키기



위한 프로세서를 포함하고 상기 인스트럭션이 실행될 때, 상기 프로세서는 복수의 동작들을 수행하고 상기 복수의 동작들은 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환하는 동작과, 상기 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화하는 동작과, 상기 주파수 영역의 오디오 신호를 상기 양자화된 에너지에 따라 정규화한 정규 신호를 생성하는 동작과, 상기 정규 신호 및 상기 입력 오디오 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터(feature vector)를 획득하는 동작과, 상기 특징 벡터를 양자화하는 동작과, 양자화된 특징 벡터에 기초하여, 상기 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득하는 동작과, 상기 스케일 값에 기초하여 상기 정규 신호를 스케일링한 조정 신호를 양자화하는 동작과, 상기 양자화된 에너지, 상기 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bit stream)들을 출력하는 동작을 포함할 수 있다.

[0024] 상기 특징 벡터를 획득하는 동작은 상기 입력 오디오 신호에 기초하여, 주파수 영역내 상기 입력 오디오 신호의 크기 스펙트럼(magnitude spectrum)을 획득하는 동작과, 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 특징 벡터를 획득하는 동작을 포함할 수 있다.

[0025] 상기 특징 벡터를 획득하는 동작은 상기 크기 스펙트럼 및 상기 정규 신호에 기초하여, 상기 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 생성하는 동작과, 상기 잠재 표현에 기초하여, 상기 특징 벡터를 계산하는 동작을 포함할 수 있다.

[0026] 상기 스케일 값을 획득하는 동작은 상기 양자화된 특징 벡터에 기초하여, 상기 주파수 밴드 별로 스케일 값을 획득하는 동작을 포함할 수 있다. 상기 스케일 값의 차원 수는 상기 주파수 밴드의 총 밴드 수와 일치하는 것일 수 있고 상기 양자화된 특징 벡터의 차원 수는 상기 특징 벡터의 차원 수와 일치하는 것일 수 있다.

[0027] 상기 조정 신호를 양자화하는 동작은 상기 정규 신호를 상기 스케일 값에 따라 스케일링하여, 상기 조정 신호를 생성하는 동작을 포함할 수 있다.

[0028] 상기 비트 스트림들을 출력하는 동작은 상기 양자화된 특징 벡터를 부호화하여 제1 비트 스트림을 출력하는 동작과, 상기 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력하는 동작과, 상기 양자화된 에너지를 부호화하여, 제3 비트 스트림을 출력하는 동작을 포함할 수 있다.

## 도면의 간단한 설명

[0030] 도 1은 일 실시예에 따른 부호화기 및 복호화기를 나타낸다.

도 2는 도 1에 도시된 부호화기의 블록도의 일 예를 나타낸다.

도 3은 도 1에 도시된 복호화기의 블록도의 일 예를 나타낸다.

도 4는 일 실시예에 따른 부호화 방법의 흐름도의 일 예를 나타낸다.

도 5는 일 실시예에 따른 복호화 방법의 흐름도의 일 예를 나타낸다.

도 6은 일 실시예에 따른 부호화기 및 복호화기를 학습하는 방법을 설명하기 위한 도면이다.

도 7은 일 실시예에 따른 장치의 일 예를 나타낸다.

## 발명을 실시하기 위한 구체적인 내용

[0031] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.

[0032] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.

[0033] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.

[0034] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다"

또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0035] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0036] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

[0038] 도 1은 일 실시예에 따른 부호화기 및 복호화기를 나타낸다.

[0039] 도 1을 참조하면, 부호화기(110)는 입력 오디오 신호를 부호화 하여 비트 스트림(bit stream)을 생성하고, 비트 스트림을 복호화기(130)로 전송(또는 출력)할 수 있다. 복호화기(130)는 부호화기(110)로부터 수신한 비트 스트림을 복호화 하여 복원 오디오 신호를 생성할 수 있다.

[0040] 부호화기(110)는 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환할 수 있다. 부호화기(110)는 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화할 수 있다. 부호화기(110)는 주파수 영역의 오디오 신호를 양자화된 에너지에 따라 정규화한 정규 신호를 생성할 수 있다. 부호화기(110)는 정규 신호 및 입력 오디오 신호에 기초하여, 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터(feature vector)를 획득할 수 있다. 부호화기(110)는 특징 벡터를 양자화할 수 있다. 부호화기(110)는 양자화된 특징 벡터에 기초하여, 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득할 수 있다. 부호화기(110)는 스케일 값에 기초하여 정규 신호를 스케일링한 조정 신호를 양자화할 수 있다. 부호화기(110)는 양자화된 에너지, 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여 비트 스트림들을 출력할 수 있다. 비트 스트림은 양자화된 특징 벡터를 부호화한 제1 비트 스트림, 양자화된 조정 신호를 부호화한 제2 비트 스트림, 및 양자화된 에너지를 부호화한 제3 비트 스트림을 포함할 수 있다.

[0041] 복호화기(130)는 부호화기(110)로부터 비트 스트림들을 수신할 수 있다. 복호화기(130)는 양자화된 특징 벡터를 부호화한 제1 비트 스트림에 기초하여, 복원 조정 신호를 역 스케일링 하는데 사용되는 스케일 값을 획득할 수 있다. 복호화기(130)는 스케일 값, 및 양자화된 조정 신호를 부호화한 제2 비트 스트림에 기초하여 복원 정규 신호를 생성할 수 있다. 복호화기(130)는 양자화된 에너지를 부호화한 제3 비트 스트림, 및 복원 정규 신호에 기초하여, 주파수 영역의 복원 오디오 신호를 획득할 수 있다. 복호화기(130)는 주파수 영역의 복원 오디오 신호에 기초하여, 시간 영역의 복원 오디오 신호를 출력할 수 있다.

[0043] 도 2는 도 1에 도시된 부호화기의 블록도의 일 예를 나타낸다.

[0044] 도 2를 참조하면, 부호화기(110)는 변형 이산 코사인 변환기(modified discrete cosine transformer)(210), 밴드 에너지 계산기(band energy calculator)(231), 양자화기(quantizer)(233, 255, 및 275), 허프만 부호화기(huffman encoder)(235, 257, 및 277), 밴드별 정규화기(band-wise normalizer)(251), 밴드별 스케일 조절기(band-wise scaler)(253), DFT 변환기(discrete fourier transformer)(271), 프레임 임베딩 추출기(frame embedding extractor)(273), 및 스케일 값 예측기(scale factor estimator)(290)를 포함할 수 있다.

[0045] 부호화기(110)는 변형 이산 코사인 변환기(210)를 통해 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환할 수 있다. 예를 들어, 변형 이산 코사인 변환기(210)는 시간 영역의 입력 오디오 신호 중 유한한 길이 L만큼을 L/2간격마다 추출할 수 있다. 변형 이산 코사인 변환기(210)는 각 오디오 신호의 프레임에 L의 길이를 가지는 윈도우 함수(window function)를 곱할 수 있다. 변형 이산 코사인 변환기(210)는 시간 영역의 입력 오디오 신호를 L/2의 길이를 갖는 주파수 영역의 오디오 신호로 변환시킬 수 있다. 시간 영역의 입력 오디오 신호의 길이 및 윈도우 함수의 선택에 의해 변형 이산 코사인 변환기(210)의 구성이 한정되는 것은 아니다. 다만, 변형 이산 코사인 변환기(210)와 역 변형 이산 코사인 변환기(예: 도 3의 370)간에 상호 역변환

관계는 유지되어야 한다.

[0046] 부호화기(110)는 밴드 에너지 계산기(231)을 통해 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 계산할 수 있다. 예를 들어, 밴드 에너지 계산기(231)는 주파수 영역의 오디오 신호의 인접한 계수들끼리 중복되지 않게 서로 묶어 주파수 영역의 오디오 신호를 복수개의 밴드로 나눌 수 있다. 밴드 에너지 계산기(231)는 수학적 식 1에 따라 각 밴드 별로 주파수 영역의 오디오 신호의 계수들이 가지는 평균 에너지를 계산할 수 있다. 밴드는 복수개의 주파수 영역의 오디오 신호의 계수를 포함할 수 있다. 하나의 밴드에 포함된 계수들은 주파수 영역에서 서로 인접해 있어야 한다. 각 밴드에 포함된 계수의 수는 서로 다를 수 있다. 다만, 하나의 계수는 복수의 밴드에 포함될 수 없다.

[0048] [수학적 식 1]

$$e_b = \sqrt{\sum_{i=i_b}^{i_{b+1}-1} y_i^2}$$

[0049]

[0051] 수학적 식 1에서  $b$ 는 밴드의 인덱스(index)를 나타내고,  $i_b$ 는  $b$ 번째 밴드에 소속된 가장 첫 번째 주파수 성분의 주파수 영역 인덱스를 나타낸다.  $e_b$ 는  $b$ 번째 밴드의 평균 에너지를 나타내고,  $y_i$ 는 주파수 영역의 오디오 신호의 계수를 나타낸다.

[0052] 부호화기(110)는 양자화기(233)을 통해 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화 할 수 있다. 예를 들어, 양자화기(233)는 균일 미드 쓰레드 양자화기(uniform mid-thread quantizer)일 수 있다. 양자화기(233)는 각 밴드 에너지를  $\log$  함수를 거친 뒤 밴드 별 고유의 양자화 단계 크기(step size)에 따라 양자화할 수 있다.

[0053] 부호화기(110)는 밴드 별 정규화기(251)을 통해 주파수 영역의 오디오 신호를 양자화된 에너지에 따라 정규화한 정규 신호를 생성할 수 있다. 예를 들어, 밴드 별 정규화기(251)는 주파수 영역의 오디오 신호를 양자화된 밴드 에너지에 의해 각 밴드 별로 정규화할 수 있다. 밴드 별 정규화기(251)는 수학적 식 2에 따라 정규 신호를 생성할 수 있다.

[0055] [수학적 식 2]

$$f_i = \frac{y_i}{\hat{e}_b}$$

[0056]

[0058] 수학적 식 2에서  $i$ 는 주파수 영역 오디오 신호의 계수의 인덱스를 나타내며,  $b$ 는 해당 계수가 속하는 밴드의 인덱스를 나타낸다.  $\hat{e}_b$ 는 양자화된 주파수 밴드의 에너지,  $y_i$ 는 주파수 영역의 오디오 신호의 계수, 및  $f_i$ 는 정규 신호의 계수를 나타낸다.

[0059] 부호화기(110)는 정규 신호 및 시간 영역의 입력 오디오 신호에 기초하여, 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터를 획득할 수 있다. 부호화기(110)는 입력 오디오 신호에 기초하여, 주파수 영역내 입력 오디오 신호의 크기 스펙트럼(magnitude spectrum)을 획득할 수 있다. 부호화기(110)는 크기 스펙트럼 및 정규 신호에 기초하여, 특징 벡터를 획득할 수 있다. 부호화기(110)는 크기 스펙트럼 및 정규 신호에 기초하여, 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 생성할 수 있다. 부호화기(110)는 잠재 표현에 기초하여 특징 벡터를 계산할 수 있다.

[0060] 예를 들어, 부호화기(110)는 DFT 변환기(271)를 통해 시간 영역의 입력 오디오 신호의 크기 스펙트럼을 획득할 수 있다. DFT 변환기(271)는 입력 오디오 신호에 윈도우 함수를 적용할 수 있다. DFT 변환기(271)는 윈도우 함수를 적용한 신호를 이산 푸리에 변환하여 크기 스펙트럼을 획득할 수 있다. 부호화기(110)는 프레임 임베딩

추출기(273)를 통해 특징 벡터를 획득할 수 있다. 프레임 임베딩 추출기(273)는 정규 신호 및 크기 스펙트럼에 기초하여, 특징 벡터를 생성할 수 있다. 프레임 임베딩 추출기(273)는 정규 신호 및 크기 스펙트럼에 기초하여 합성곱 신경망(convolutional neural network)을 통해 주파수 밴드의 에너지에 관한 정보를 추출하기 위한 잠재 표현(latent representation)을 획득할 수 있다. 프레임 임베딩 추출기(273)는 다계층 퍼셉트론(multi-layer perceptron, MLP)을 통해 잠재 표현으로부터 특징 벡터를 계산할 수 있다. 특징 벡터는 주파수 영역의 오디오 신호의 주파수 밴드의 에너지에 관한 정보를 포함할 수 있다. 특징 벡터는 주파수 영역 에너지 분포로부터 얻어진 것일 수 있다.

[0061] 부호화기(110)는 특징 벡터를 양자화할 수 있다. 예를 들어, 부호화기(110)는 양자화기(275)를 통해 특징 벡터를 양자화할 수 있다. 양자화기(275)는 균일 미드 쓰레드 양자화기일 수 있다. 양자화기(275)는 복수의 차원을 갖는 특징 벡터에 대하여 각 차원 별로 양자화 단계 크기를 할당할 수 있다. 양자화기(275)는 특징 벡터의 각 차원 별로 양자화 할 수 있다.

[0062] 부호화기(110)는 양자화된 특징 벡터에 기초하여, 정규 신호를 스케일링 하는데 사용되는 스케일 값(scale factor)을 획득할 수 있다. 부호화기(110)는 양자화된 특징 벡터에 기초하여, 주파수 밴드 별로 스케일 값을 획득할 수 있다. 스케일 값의 차원 수는 주파수 밴드의 총 밴드 수와 일치할 수 있다. 양자화된 특징 벡터의 차원 수는 특징 벡터의 차원 수와 일치할 수 있다.

[0063] 예를 들어, 부호화기(110)는 스케일 값 예측기(290)을 통해 스케일 값을 획득할 수 있다. 스케일 값 예측기(290)는 복수의 층을 가진 다계층 퍼셉트론으로 구성될 수 있으나, 이에 한정되는 것은 아니다. 스케일 값 예측기(290)는 양자화된 특징 벡터에 기초하여 각 주파수 밴드의 스케일 값을 계산할 수 있다. 스케일 값 예측기(290)는 양수의 스케일 값을 획득하기 위해 지수함수를 사용할 수 있다.

[0064] 부호화기(110)는 스케일 값에 기초하여, 정규 신호를 스케일링한 조정 신호를 양자화할 수 있다. 부호화기(110)는 정규 신호를 스케일 값에 따라 스케일링 하여, 조정 신호를 생성할 수 있다. 예를 들어, 부호화기(110)는 밴드별 스케일 조절기(253)를 통해 조정 신호를 생성할 수 있다. 밴드별 스케일 조절기(253)는 정규 신호를 수학식 3에 따라 스케일링하여 조정 신호를 생성할 수 있다. 같은 밴드 내에 포함된 정규 신호들은 동일한 스케일 값에 의해 스케일링 될 수 있다. 또한, 양자화기(255)는 조정 신호를 양자화할 수 있다. 양자화기(255)는 균일 미드 쓰레드 양자화기일 수 있다. 양자화기(255)는 주파수 밴드 별로 양자화 단계 크기를 할당할 수 있다. 동일한 주파수 밴드에 포함된 정규 신호들은 동일한 양자화 단계 크기에 의해 양자화될 수 있다.

[0066] [수학식 3]

$$f_{s,i} = \sigma_b f_i$$

[0069] 수학식 3에서,  $i$ 는 정규 신호의 계수의 인덱스를 나타내고,  $b$ 는 해당 계수가 속하는 밴드의 인덱스를 나타낸다.  $\sigma_b$ 는  $b$ 번째 밴드의 스케일 값을 나타내고,  $f_i$  정규 신호의 계수를 나타내고,  $f_{s,i}$ 는 조정 신호의 계수를 나타낸다.

[0070] 부호화기(110)는 양자화된 에너지, 양자화된 특징 벡터, 및 양자화된 조정 신호에 기초하여, 비트 스트림(bitstream)들을 출력할 수 있다. 부호화기(110)는 양자화된 특징 벡터를 부호화 하여 제1 비트 스트림을 출력할 수 있다. 부호화기(110)는 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력할 수 있다. 부호화기(110)는 양자화된 에너지를 부호화하여 제3 비트 스트림을 출력할 수 있다. 부호화기(110)는 허프만 부호화기(예: 235, 257, 및 277)을 통해 비트 스트림을 출력할 수 있다. 예를 들어, 허프만 부호화기(277)는 양자화된 특징 벡터를 부호화하여 제1 비트 스트림을 출력할 수 있다. 허프만 부호화기(277)는 복수의 차원을 갖는 특징 벡터에 대해 각 차원별로 코드북을 할당할 수 있다. 허프만 부호화기(257)는 양자화된 조정 신호를 부호화하여 제2 비트 스트림을 출력할 수 있다. 허프만 부호화기(235)는 양자화된 에너지를 부호화하여 제3 비트 스트림을 출력할 수 있다. 허프만 부호화기(235, 및 257)는 주파수 밴드별로 동일한 코드북을 할당할 수 있다.

[0071] 허프만 부호화기(235, 257, 및 277)는 주파수 밴드별로 공백 여부를 판정하는 비트를 생성할 수 있다. 예를 들어, 허프만 부호화기(235, 257, 및 277)는 주파수 밴드의 에너지가 사용자가 설정한 최소 임계값 이하인 경우, 최소 임계값 이하인 주파수 밴드에 대응되는 비트를 1로 설정할 수 있다. 허프만 부호화기(235, 257, 및 277)

는 주파수 밴드의 에너지가 일정 수준 이하인 주파수 밴드내에 포함되는 조정 신호를 저장 혹은 전송하지 않을 수 있다.

[0072] 또한, 허프만 부호화기(235, 257, 및 277)는 비트 스트림 생성시 사용자가 설정한 최대 주파수 밴드의 수에 대응되는 정보를 수신할 수 있다. 허프만 부호화기(235, 257, 및 277)는 지정된 주파수 밴드에 대해서만 신호를 부호화할 수 있다.

[0074] 도 3은 도 1에 도시된 복호화기의 블록도의 일 예를 나타낸다.

[0075] 도 3을 참조하면, 복호화기(130)는 허프만 복호화기(huffman decoder)(311, 331, 및 351), 스케일값 예측기(313), 밴드 별 역 스케일 조절기(band-wise inverse scaler)(333), 밴드별 역 정규화기(band-wise denormalizer)(353), 및 역 변형 이산 코사인 변환기(inverse modified discrete cosine transformer)(370)를 포함할 수 있다.

[0076] 복호화기(130)는 양자화된 특징 벡터를 부호화한 제1 비트 스트림에 기초하여, 복원 조정 신호를 역 스케일링 하는데 사용되는 스케일 값을 획득할 수 있다. 복호화기(130)는 제1 비트 스트림을 복호화 하여 양자화된 특징 벡터를 획득할 수 있다. 복호화기(130)는 양자화된 특징 벡터로부터 스케일 값을 계산할 수 있다. 예를 들어, 복호화기(130)는 허프만 복호화기(311)을 통해 제1 비트 스트림을 복호화하여, 양자화된 특징 벡터를 획득할 수 있다. 허프만 복호화기(311)는 허프만 부호화기(277)에서 사용된 코드북과 동일한 코드북을 사용하여 제1 비트 스트림을 복호화할 수 있다. 복호화기(130)는 스케일 값 예측기(313)을 통해 양자화된 특징 벡터로부터 스케일 값을 계산할 수 있다. 복호화기(130)에서의 스케일 값 예측기(313)은 부호화기(110)에서의 스케일 값 예측기(290)와 동일한 구성을 가질 수 있다. 스케일 값 예측기(313)에서 계산된 스케일 값은 스케일 값 예측기(290)에서 계산된 스케일 값과 동일한 값을 가질 수 있다.

[0077] 복호화기(130)는 스케일 값, 및 양자화된 조정 신호를 부호화한 제2 비트 스트림에 기초하여 복원 정규 신호를 생성할 수 있다. 복호화기(130)는 제2 비트 스트림을 복호화 하여 복원 조정 신호를 생성할 수 있다. 복호화기(130)는 복원 조정 신호를 스케일 값에 따라 역 스케일링할 수 있다.

[0078] 예를 들어, 복호화기(130)는 허프만 복호화기(331)을 통해 제2 비트 스트림을 복호화하여 복원 조정 신호를 생성할 수 있다. 허프만 복호화기(331)는 허프만 부호화기(257)에서 사용된 코드북과 동일한 코드북을 사용하여 제2 비트 스트림을 복호화할 수 있다. 복호화기(130)는 밴드 별 역 스케일 조절기(333)을 통해 복원 조정 신호를 스케일 값에 따라 역 스케일링 할 수 있다. 밴드 별 역 스케일 조절기(333)는 수학적 식 4에 따라 복원 조정 신호를 역 스케일링하여 복원 정규 신호를 생성할 수 있다.

[0080] [수학적 식 4]

$$[0081] \hat{f}_i = \hat{f}_{si} / \sigma_b$$

[0083] 수학적 식 4에서,  $i$ 는 복원 조정 신호의 계수의 인덱스를 나타내고,  $b$ 는 해당 신호가 속하는 밴드의 인덱스를 나타낸다.  $\sigma_b$ 는  $b$ 번째 밴드의 스케일 값을 나타내고,  $\hat{f}_i$  복원 정규 신호의 계수를 나타내고,  $\hat{f}_{si}$ 는 복원 조정 신호의 계수를 나타낸다.

[0084] 복호화기(130)는 양자화된 에너지를 부호화한 제3 비트 스트림, 및 복원 정규 신호에 기초하여, 주파수 영역의 복원 오디오 신호를 획득할 수 있다. 복호화기(130)는 제3 비트 스트림에 기초하여 주파수 영역의 복원 오디오 신호의 주파수 밴드의 복원 에너지를 출력할 수 있다. 복호화기(130)는 복원 정규 신호를 복원 에너지에 따라 역 정규화할 수 있다.

[0085] 예를 들어, 복호화기(130)는 허프만 복호화기(351)을 통해 제3 비트 스트림을 복호화하여 주파수 밴드의 복원 에너지를 출력할 수 있다. 허프만 복호화기(351)는 허프만 부호화기(235)에서 사용된 코드북과 동일한 코드북을 사용하여 제3 비트 스트림을 복호화할 수 있다. 복호화기(130)는 밴드 별 역 정규화기(353)을 통해 복원 정규 신호를 역 정규화할 수 있다. 밴드 별 역 정규화기(353)는 수학적 식 5에 따라서 복원 정규 신호를 역 정규화



할 수 있다.

[0087] [수학식 5]

$$[0088] \hat{y}_i = \hat{e}_b \hat{f}_i$$

[0090] 수학식 5에서,  $i$ 는 복원 정규 신호의 계수의 인덱스를 나타내고,  $b$ 는 해당 신호가 속하는 밴드의 인덱스를 나타낸다.  $\hat{e}_b$ 는  $b$ 번째 밴드의 복원 에너지를 나타내고,  $\hat{f}_i$  복원 정규 신호의 계수를 나타내고,  $\hat{y}_i$ 는 주파수 영역의 복원 오디오 신호의 계수를 나타낸다.

[0091] 복호화기(130)는 주파수 영역의 복원 오디오 신호에 기초하여, 시간 영역의 복원 오디오 신호를 출력할 수 있다. 복호화기(130)는 역 변형 이산 코사인 변환기(370)을 통해 시간 영역의 복원 오디오 신호를 출력할 수 있다. 예를 들어, 역 변형 이산 코사인 변환기(370)는 주파수 영역의 복원 오디오 신호를 역 변형 이산 코사인 변환하여 길이가  $L$ 인 시간 영역 신호로 변환시킬 수 있다. 역 변형 이산 코사인 변환기(370)는 길이가  $L$ 인 시간 영역 신호에 윈도우 함수를 적용할 수 있다. 역 변형 이산 코사인 변환기(370)는 윈도우 함수가 적용된 이웃한 신호들에 중첩 및 가산(overlap - and add)연산을 수행하여 시간 영역의 복원 오디오 신호를 출력할 수 있다.

[0092] 허프만 복호화기(311, 331, 및 351)는 도 2의 허프만 부호화기(235, 257, 및 277)에서 생성된 주파수 밴드별로 공백 여부를 판정하는 비트를 수신할 수 있다. 예를 들어, 허프만 복호화기(311, 331, 및 351)는 공백 여부를 판정하는 비트 값이 1인 경우 모든 값을 0으로 복원할 수 있다. 허프만 복호화기(311, 331, 및 351)는 수신한 주파수 밴드의 에너지를 최소 임계값과 비교할 수 있다. 허프만 복호화기(311, 331, 및 351)는 수신한 주파수 밴드의 에너지가 최소 임계값 보다 낮은 주파수 밴드는 비어 있는 것으로 판단할 수 있다.

[0094] 도 4는 일 실시예에 따른 부호화 방법의 흐름도의 일 예를 나타낸다.

[0095] 동작 405 내지 동작 440은 순차적으로 수행될 수 있지만, 이에 한정되는 것은 아니다. 예를 들어, 둘 이상의 동작들이 병렬적으로 수행될 수도 있다.

[0096] 동작 405에서, 부호화기(110)는 시간 영역의 입력 오디오 신호를 주파수 영역의 오디오 신호로 변환시킬 수 있다.

[0097] 동작 410에서, 부호화기(110)는 주파수 영역의 오디오 신호의 주파수 밴드의 에너지를 양자화할 수 있다.

[0098] 동작 415에서, 부호화기(110)는 주파수 영역의 오디오 신호를 양자화된 에너지에 따라 정규화한 정규 신호를 생성할 수 있다.

[0099] 동작 420에서, 부호화기(110)는 정규 신호 및 입력 오디오 신호에 기초하여, 주파수 밴드의 에너지에 관한 정보를 포함하는 특징 벡터를 획득할 수 있다.

[0100] 동작 425에서, 부호화기(110)는 특징 벡터를 양자화할 수 있다.

[0101] 동작 430에서, 부호화기(110)는 양자화된 특징 벡터에 기초하여, 정규 신호를 스케일링 하는데 사용되는 스케일 값을 획득할 수 있다.

[0102] 동작 435에서, 부호화기(110)는 스케일 값에 기초하여 정규 신호를 스케일링한 조정 신호를 양자화할 수 있다.

[0103] 동작 440에서, 부호화기(110)는 양자화된 에너지, 양자화된 특징 벡터 및 양자화된 조정 신호에 기초하여 비트 스트림을 출력할 수 있다.

[0105] 도 5는 일 실시예에 따른 복호화 방법의 흐름도의 일 예를 나타낸다.

[0106] 동작 510 내지 동작 590은 순차적으로 수행될 수 있지만, 이에 한정되는 것은 아니다. 예를 들어, 둘 이상의

동작들이 병렬적으로 수행될 수도 있다.

[0107] 동작 510에서, 복호화기(130)는 부호화기(110)로부터 비트 스트림들을 수신할 수 있다.

[0108] 동작 530에서, 복호화기(130)는 양자화된 특징 벡터를 부호화한 제1 비트 스트림에 기초하여, 복원 조정 신호를 역 스케일링 하는데 사용되는 스케일 값을 획득할 수 있다.

[0109] 동작 550에서, 복호화기(130)는 스케일 값, 및 양자화된 조정 신호를 부호화한 제2 비트 스트림에 기초하여, 복원 정규 신호를 생성할 수 있다.

[0110] 동작 570에서, 복호화기(130)는 양자화된 에너지를 부호화한 제3 비트 스트림, 및 복원 정규 신호에 기초하여 주파수 영역의 복원 오디오 신호를 획득할 수 있다.

[0111] 동작 590에서, 복호화기(130)는 주파수 영역의 복원 오디오 신호에 기초하여, 시간 영역의 복원 오디오 신호를 출력할 수 있다.

[0113] 도 6은 일 실시예에 따른 부호화기 및 복호화기를 학습하는 방법을 설명하기 위한 도면이다.

[0114] 학습 장치(600)는 도 1의 부호화기 및 복호화기를 하기의 학습과정을 통해 학습시킬 수 있다.

[0115] 학습을 위한 손실 함수는 수학식 6과 같이 왜곡  $\mathcal{D}$ 와 비트 전송률  $\mathcal{R}$ 의 가중합으로 구성될 수 있다.

[0117] [수학식 6]

$$\mathcal{L} = \mathcal{R} + \lambda \mathcal{D}$$

[0120] 수학식 6에서  $\mathcal{L}$ 은 부호화기(110) 및 복호화기(130)의 학습을 위한 손실 함수를 나타내고,  $\mathcal{D}$ 는 왜곡을 나타내고,  $\mathcal{R}$ 은 비트 전송률을 나타낸다.

[0121] 비트 전송률  $\mathcal{R}$ 은 엔트로피 모델(entropy model)을 통해 계산된 각 변수의 엔트로피로 결정될 수 있다. 각 변수의 엔트로피는 엔트로피 코딩(entropy coding)을 통해 실제 비트율 생성시 달성 가능한 평균 비트 스트림 길이의 하한선일 수 있다. 엔트로피 모델은 특정 변수의 확률 밀도를 제공할 수 있다. 엔트로피 모델은 입력 변수의 확률 분포를 학습함과 동시에 입력 변수의 엔트로피 코딩시의 비트 스트림의 길이를 추산하여 비트 전송률  $\mathcal{R}$ 을 최적화할 수 있다.

[0122] 왜곡  $\mathcal{D}$ 는 수학식 7에 따라 계산될 수 있으나, 이에 한정되는 것은 아니다. 학습장치(600)는 입력 오디오 신호에 대하여 주파수 영역에서 마스킹 효과가 발생하는 임계점(masking threshold)을 미리 계산할 수 있다. 학습장치(600)는 입력 오디오 신호와 최종 복원 오디오 신호의 차이로부터 왜곡 신호를 얻어낼 수 있다. 학습장치(600)는 왜곡 신호를 주파수 영역 신호로 변환하여 왜곡 신호가 마스킹 효과 임계점보다 큰 양의 총합을 왜곡  $\mathcal{D}$ 로 계산한다. 마스킹 효과 임계점은 심리 음향 모델(psychoacoustic model)을 통해 계산될 수 있으나, 이에 한정되는 것은 아니다. 학습장치(600)는 심리 음향 모델에 따라 마스킹 효과 임계점의 크기를 구할 수 있다. 다만, 심리 음향 모델은 학습 과정에서만 사용될 뿐, 실제 부호화기(110) 및 복호화기(130)의 이용시 사용되지 않을 수 있다.

[0124] [수학식 7]

$$\mathcal{D} = \frac{1}{N} \sum_{i=1}^N \max(\text{SPL}_i(\mathbf{x} - \hat{\mathbf{x}}) - M_i(\mathbf{x}), 0)$$

[0125]

- [0127] 수학식 7에서,  $x$ 와  $\hat{x}$ 는 각각 입력 오디오 신호와 최종 복원된 오디오 신호를 나타내고,  $SPL_i(x - \hat{x})$ 는 왜곡 신호  $x - \hat{x}$ 를 주파수 영역으로 변환된 뒤 음압 수준(sound pressure level)을 계산하였을 때  $i$ 번째 주파수 성분이 가지는 음압의 크기를 나타내고,  $M_i(x)$ 는 입력 오디오 신호에 대하여 계산된 마스킹 효과 임계점을 나타내고,  $N$ 은 총 주파수 성분의 수를 나타낸다.
- [0129] 도 7은 일 실시예에 따른 장치의 일 예를 나타낸다.
- [0130] 도 7을 참조하면, 장치(700)는 메모리(710) 및 프로세서(730)를 포함할 수 있다. 장치(700)는 도 1의 부호화기(110) 또는 복호화기(130)를 포함할 수 있다. 장치(700)는 도 1의 부호화기(110) 및 복호화기(130)를 모두 포함한 장치일 수 있다.
- [0131] 메모리(710)는 프로세서(730)에 의해 실행가능한 인스트럭션들(예: 프로그램)을 저장할 수 있다. 예를 들어, 인스트럭션들은 프로세서(730)의 동작 및/또는 프로세서(730)의 각 구성의 동작을 실행하기 위한 인스트럭션들을 포함할 수 있다.
- [0132] 프로세서(730)는 메모리(710)에 저장된 데이터를 처리할 수 있다. 프로세서(730)는 메모리(710)에 저장된 컴퓨터로 읽을 수 있는 코드(예를 들어, 소프트웨어) 및 프로세서(730)에 의해 유발된 인스트럭션(instruction)들을 실행할 수 있다.
- [0133] 프로세서(730)는 목적하는 동작들(desired operations)을 실행시키기 위한 물리적인 구조를 갖는 회로를 가지는 하드웨어로 구현된 데이터 처리 장치일 수 있다. 예를 들어, 목적하는 동작들은 프로그램에 포함된 코드(code) 또는 인스트럭션들(instructions)을 포함할 수 있다.
- [0134] 예를 들어, 하드웨어로 구현된 데이터 처리 장치는 마이크로프로세서(microprocessor), 중앙 처리 장치(central processing unit), 프로세서 코어(processor core), 멀티-코어 프로세서(multi-core processor), 멀티프로세서(multiprocessor), ASIC(Application-Specific Integrated Circuit), FPGA(Field Programmable Gate Array)를 포함할 수 있다.
- [0135] 도 1의 부호화기(110), 복호화기(130), 및/또는 도 6의 학습 장치(600)는 메모리(710)에 저장되어 프로세서(730)에 의해서 실행되거나 프로세서(730)에 임베디드될 수 있다. 프로세서(730)는 도 1 내지 도 5를 참조하여 부호화기(110) 및/또는 복호화기(130)의 동작을 실질적으로 동일하게 수행할 수 있다. 또한, 프로세서(730)는 도 6을 참조하여, 학습 장치(600)의 동작을 실질적으로 동일하게 수행할 수 있다. 이에 상세한 설명은 생략하도록 한다.
- [0137] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0138] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로,



또는 일시적으로 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0139] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 저장할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

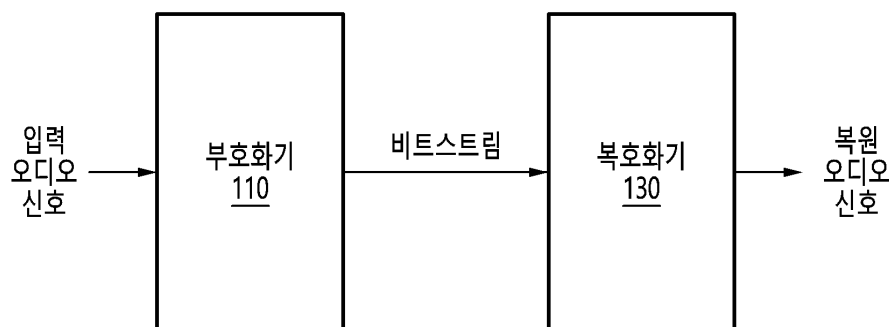
[0140] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0141] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

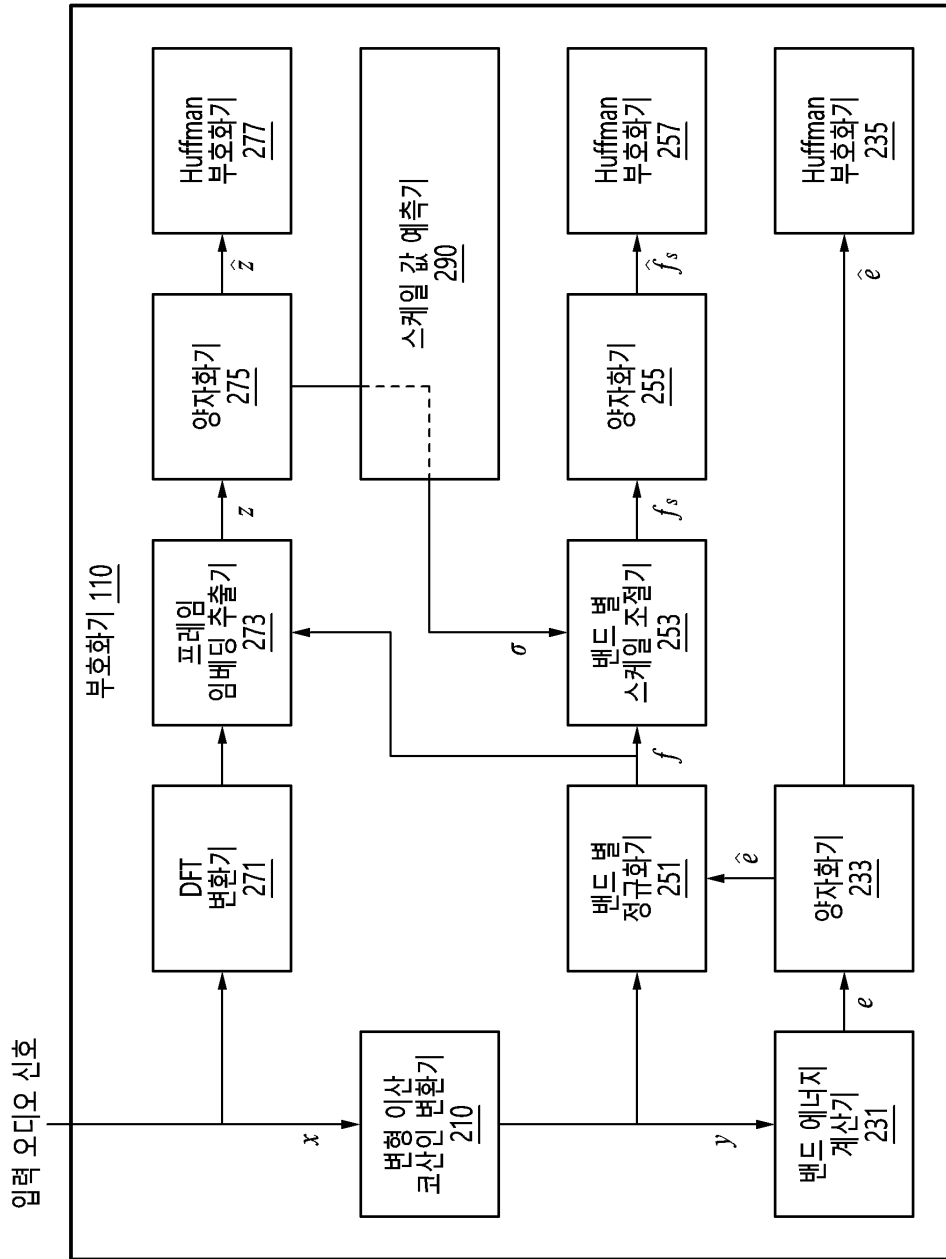
[0142] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

## 도면

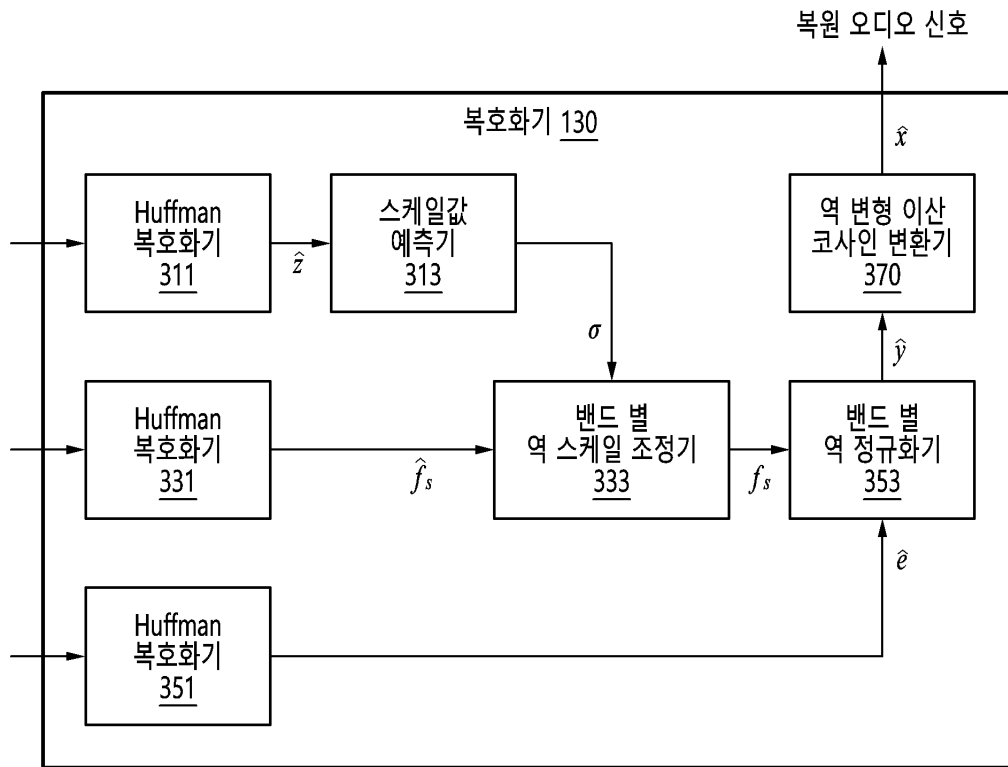
### 도면1



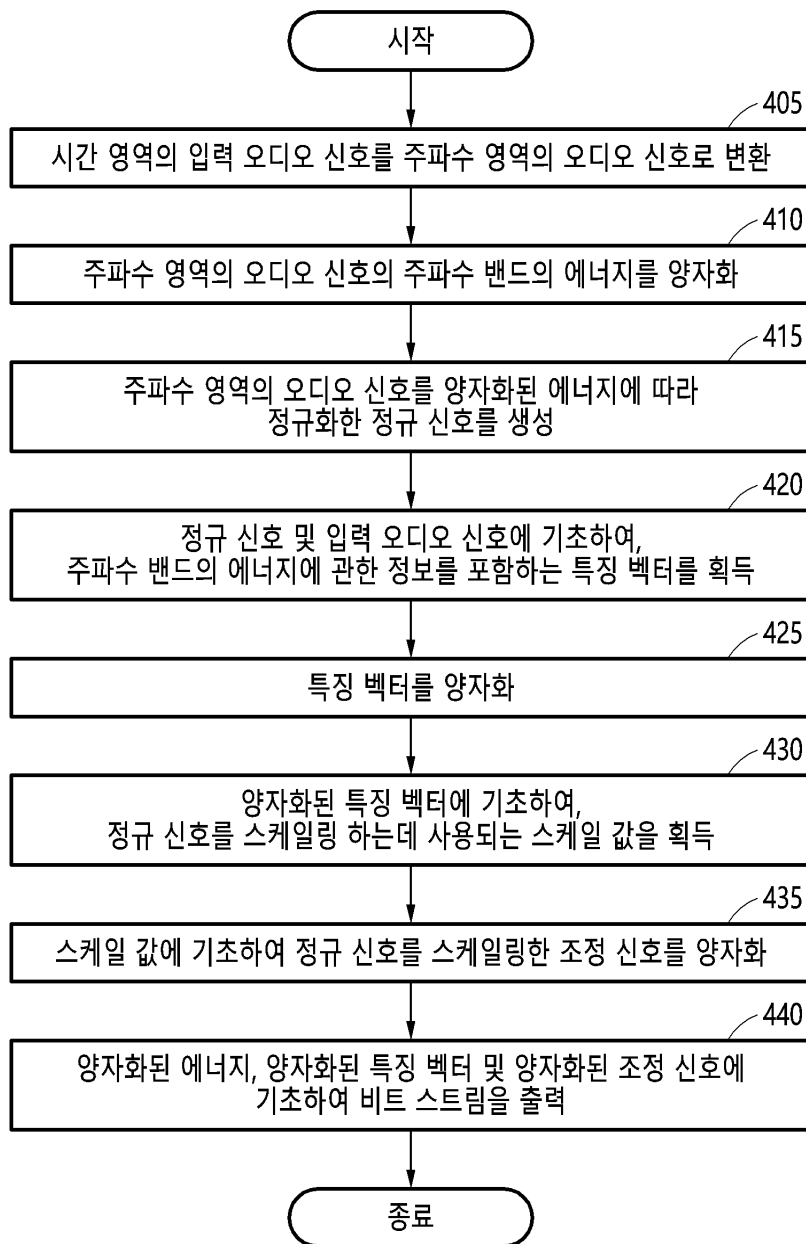
도면2



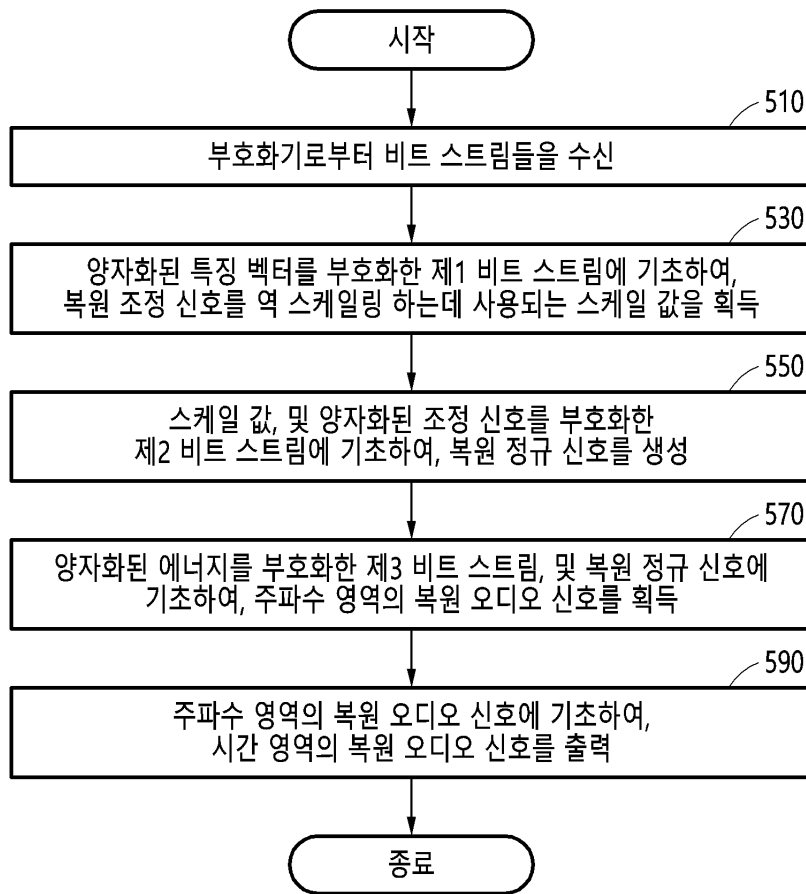
도면3



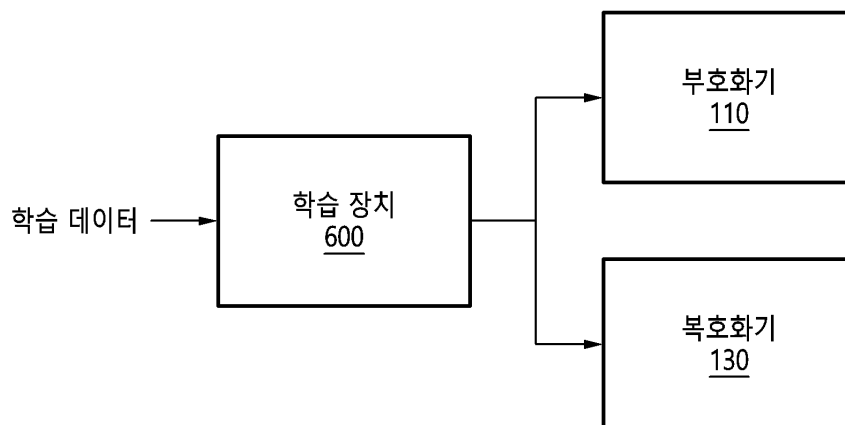
도면4



도면5



도면6



도면7

