



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0106902
(43) 공개일자 2019년09월18일

(51) 국제특허분류(Int. Cl.)

G10L 15/06 (2006.01) G10L 15/02 (2006.01)

G10L 21/02 (2006.01) G10L 25/18 (2013.01)

(52) CPC특허분류

G10L 15/063 (2013.01)

G10L 15/02 (2013.01)

(21) 출원번호 10-2019-0106666

(22) 출원일자 2019년08월29일

심사청구일자 없음

(71) 출원인

엘지전자 주식회사

서울특별시 영등포구 여의대로 128 (여의도동)

(72) 발명자

김다해

경기도 안양시 만안구 삼봉로 61, 105동 507호 (박달동, 금호타운)

(74) 대리인

특허법인 남앤남

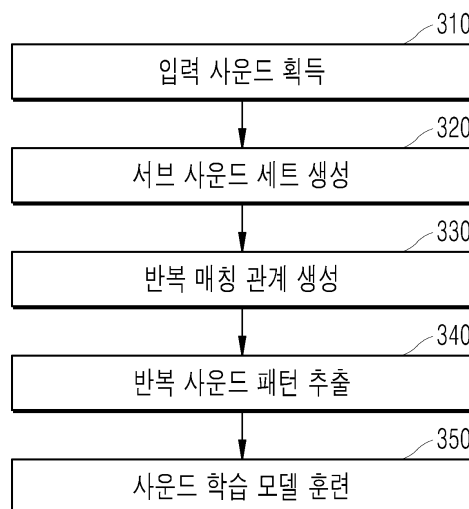
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 사운드 분석 방법 및 장치

(57) 요약

탑재된 인공지능(artificial intelligence, AI) 알고리즘 및/또는 기계학습(machine learning) 알고리즘을 실행하며, 5G 통신 환경에서 다른 전자 기기들 및 서버와 통신할 수 있는 사운드 분석 방법 및 장치가 개시된다. 사운드 분석 방법 및 장치는 타겟 공간의 사운드 환경에 특화된 사운드 학습 모델을 제공한다.

대표도 - 도3



(52) CPC특허분류

G10L 21/02 (2013.01)

G10L 25/18 (2013.01)

명세서

청구범위

청구항 1

사운드 분석 방법으로서,

타겟 공간에서 발생한 입력 사운드를 획득하는 단계;

상기 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하는 단계;

상기 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하는 단계;

상기 반복 매칭 관계에 기반하여 상기 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하는 단계;
및

상기 반복 사운드 패턴에 기반하여 상기 타겟 공간에 대한 사운드 학습 모델을 훈련하는 단계

를 포함하는 사운드 분석 방법.

청구항 2

제 1 항에 있어서,

상기 입력 사운드를 획득하는 단계는,

상기 타겟 공간에서 소정의 시간 주기마다 상기 입력 사운드를 획득하는,

를 포함하는 사운드 분석 방법.

청구항 3

제 1 항에 있어서,

상기 일 세트의 서브 사운드의 각 서브 사운드를 상기 소정의 시간 간격과는 상이한 또다른 소정의 시간 간격에 따라 반복적으로 분할하여 또다른 세트의 서브 사운드를 생성하는 단계

를 더 포함하는 사운드 분석 방법.

청구항 4

제 1 항에 있어서,

상기 일 세트의 서브 사운드를 생성한 분할과는 상이한 시작 지점부터 상기 입력 사운드를 상기 시간 간격에 따라 분할하여 또다른 세트의 서브 사운드를 생성하는 단계

를 더 포함하는, 사운드 분석 방법.

청구항 5

제 1 항에 있어서,

상기 반복 매칭 관계를 결정하는 단계는,

상기 일 세트의 서브 사운드의 하나의 서브 사운드를 타겟 서브 사운드로 선택하는 단계; 및

상기 일 세트의 서브 사운드의 나머지 서브 사운드 중 상기 타겟 서브 사운드와 파형이 동일한 서브 사운드를 상기 타겟 서브 사운드에 대한 반복 매칭 관계에 추가하는 단계를 포함하는 사운드 분석 방법.

청구항 6

제 1 항에 있어서,
상기 반복 매칭 관계를 결정하는 단계는,
상기 일 세트의 서브 사운드의 하나의 서브 사운드를 타겟 서브 사운드로 선택하는 단계; 및
또다른 세트의 서브 사운드 중 상기 타겟 서브 사운드와 파형이 동일한 서브 사운드를 상기 타겟 서브 사운드에 대한 반복 매칭 관계에 추가하는 단계를 포함하고,
상기 또다른 세트의 서브 사운드는 상기 일 세트의 서브 사운드와 동일한 타겟 장소에서 획득된 또다른 입력 사운드를 분할하여 생성한 것인 사운드 분석 방법.

청구항 7

제 1 항에 있어서,
상기 반복 사운드 패턴을 추출하는 단계는,
상기 일 세트의 서브 사운드에서 상기 반복 매칭 관계를 가지는 타겟 서브 사운드를 선택하는 단계;
상기 일 세트의 서브 사운드에서 상기 타겟 서브 사운드의 제 1 반복 주기를 결정하는 단계;
상기 타겟 공간에서 이전에 획득된 또다른 세트의 서브 사운드의 반복 매칭 관계에 기반하여 상기 또다른 세트의 서브 사운드에서 상기 타겟 서브 사운드의 제 2 반복 주기를 결정하는 단계; 및
상기 제 1 반복 주기와 상기 제 2 반복 주기의 차이가 임계값 이하인 경우에 상기 타겟 서브 사운드를 상기 반복 사운드 패턴으로 추출하는 단계를 포함하는 사운드 분석 방법.

청구항 8

제 1 항에 있어서,
상기 반복 사운드 패턴을 추출하는 단계는,
상기 일 세트의 서브 사운드에서 상기 반복 매칭 관계가 없는 타겟 서브 사운드를 선택하는 단계;
상기 타겟 공간에서 이전에 획득된 또다른 세트의 서브 사운드에서 상기 타겟 서브 사운드와 매칭되는 서브 사운드를 찾는 단계; 및
상기 타겟 서브 사운드의 타임 슬롯과 상기 매칭되는 서브 사운드의 타임 슬롯의 차이가 임계값 이하인 경우에 상기 타겟 서브 사운드를 상기 반복 사운드 패턴으로 추출하는 단계를 포함하는 사운드 분석 방법.

청구항 9

제 1 항에 있어서,

상기 사운드 학습 모델을 훈련하는 단계는,

상기 타겟 공간에서 추출한 상기 반복 사운드 패턴을 상기 사운드 학습 모델에 학습 데이터로 제공하는 단계를 포함하는 사운드 분석 방법.

청구항 10

제 1 항에 있어서,

상기 사운드 학습 모델을 이용하여 상기 타겟 공간에서 획득한 타겟 사운드로부터 상기 타겟 사운드의 반복 사운드 패턴을 결정하는 단계

를 더 포함하는 사운드 분석 방법.

청구항 11

제 10 항에 있어서,

상기 타겟 사운드의 반복 사운드 패턴을 결정하는 단계는,

상기 타겟 공간에서 상기 타겟 사운드를 획득하는 단계;

상기 타겟 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 타겟 서브 사운드를 생성하는 단계; 및

상기 사운드 학습 모델에 기반하여 상기 일 세트의 타겟 서브 사운드로부터 상기 타겟 사운드의 반복 사운드 패턴을 결정하는 단계

를 포함하는 사운드 분석 방법.

청구항 12

타겟 공간에서 발생한 타겟 사운드를 획득하는 마이크로폰;

사운드 학습 모델을 저장하는 메모리; 및

상기 사운드 학습 모델에 기반하여 상기 타겟 사운드 내의 반복 사운드 패턴을 결정하도록 동작하는 프로세서를 포함하고,

상기 프로세서는,

상기 타겟 공간에서 발생한 입력 사운드를 획득하고,

상기 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하고,

상기 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하고,

상기 반복 매칭 관계에 기반하여 상기 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하도록 설정되는

사운드 분석 장치.

청구항 13

제 12 항에 있어서,

상기 프로세서는,

추출된 상기 반복 사운드 패턴에 기반하여 상기 사운드 학습 모델을 훈련시키도록 더 설정되는

사운드 분석 장치.

청구항 14

제 12 항에 있어서,

상기 프로세서는,

상기 타겟 공간에서 상기 타겟 사운드를 획득하고,

상기 타겟 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 타겟 서브 사운드를 생성하고,

상기 사운드 학습 모델에 기반하여 각 타겟 서브 사운드가 상기 타겟 공간의 반복 사운드 패턴인 지 여부를 결정하도록 더 설정되는

사운드 분석 장치.

청구항 15

제 14 항에 있어서,

상기 프로세서는,

상기 타겟 공간의 반복 사운드 패턴인 것으로 결정된 타겟 서브 사운드를 상기 타겟 공간의 노이즈로 결정하도록 더 설정되는

사운드 분석 장치.

발명의 설명

기술 분야

[0001] 본 발명은 사운드 분석 방법 및 장치에 관한 것으로 보다 상세하게는 타겟 공간의 반복 사운드 패턴을 추출하고 이를 학습하는 사운드 분석 방법 및 장치에 관한 것이다.

배경 기술

[0002] 최근 인공 지능(AI) 기술과 사물인터넷(IoT) 기술의 본격적인 확산과 맞물려 음성 인식 기술의 응용 분야가 점차 확대되면서 다양한 환경에서 음성 인식 시스템이 적용되고 있다.

[0003] 공간의 사운드 환경은 음성 인식 성능에 영향을 미친다. 예를 들어, 공간의 사운드 환경이 바뀌면, 주변 잡음의 크기, 종류 및 음성 파형이 달라질 수 있다. 따라서, 음성 인식 성능을 높이기 위해서는 음성 인식 시스템이 적용될 사운드 환경에 적합한 음향 모델이 필요하다.

[0004] 선행 기술 1은 잡음 환경에서 음성 처리 장치를 지향하는 사용자의 신호를 강화하는 음성 처리 장치 및 방법을 개시한다. 선행 기술 1은 복수 개의 음원으로부터 발생한 혼합 신호로부터 미리 설정된 음원 방향을 제외한 나머지 음원 방향에서 발생한 신호를 제거함으로써 대상 신호만을 강화한다.

[0005] 선행 기술 2는 차량 내 소음 환경에서 잡음을 제거하여 음성 인식률을 높이는 음성 전처리 장치를 개시한다. 선행 기술 2는 차량 내 소음 환경의 변화에 따라 설정된 반향 파라미터를 토대로 잡음 제거 알고리즘의 파라미터를 갱신한다.

[0006] 하지만, 선행 기술 1 및 선행 기술 2는 단일 잡음 모델을 이용하므로 공간의 사운드 환경에 따른 노이즈의 특성 및 유형의 차이를 고려하지 못한다.

선행기술문헌

특허문헌

- [0007] (특허문헌 0001) 선행 기술 1: 한국공개특허공보 KR 10-2012-0102306 A
(특허문헌 0002) 선행 기술 2: 한국등록특허공보 KR 10-1592425 B1

발명의 내용

해결하려는 과제

- [0008] 본 발명의 일 과제는 음성 인식 서비스에서 공간의 사운드 환경에 따른 사운드 특성 및 유형의 차이를 고려하지 못함으로 인해 음성 인식 오류가 발생할 수 있는 종래 기술의 문제점을 해결하는 것이다.
- [0009] 본 발명의 일 과제는 타겟 공간의 사운드 환경에 특화된 사운드 학습 모델을 제공하는 것이다.
- [0010] 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0011] 상기 과제를 달성하기 위하여, 본 발명의 일 실시예에 따른 사운드 분석 방법은, 타겟 공간에서 획득한 입력 사운드로부터 반복 사운드 패턴을 추출 및 학습하여 타겟 공간에 맞는 사운드 학습 모델을 제공한다.
- [0012] 구체적으로, 사운드 분석 방법은 타겟 공간에서 발생한 입력 사운드를 획득하는 단계, 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하는 단계, 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하는 단계, 반복 매칭 관계에 기반하여 상기 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하는 단계 및 반복 사운드 패턴에 기반하여 타겟 공간에 대한 사운드 학습 모델을 훈련하는 단계를 포함할 수 있다.
- [0013] 상기 과제를 달성하기 위하여, 본 발명의 일 실시예에 따른 사운드 분석 장치는, 타겟 공간에서 발생한 타겟 사운드를 획득하는 마이크로폰, 사운드 학습 모델을 저장하는 메모리 및 사운드 학습 모델에 기반하여 타겟 사운드 내의 반복 사운드 패턴을 결정하도록 동작하는 프로세서를 포함할 수 있다.
- [0014] 구체적으로 프로세서는 타겟 공간에서 발생한 입력 사운드를 획득하고, 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하고, 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하고, 반복 매칭 관계에 기반하여 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하도록 설정될 수 있다.
- [0015] 본 발명에서 이루고자 하는 기술적 과제들의 해결 수단은 이상에서 언급한 해결 수단들로 제한되지 않으며, 언급하지 않은 또 다른 해결 수단들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

발명의 효과

- [0016] 본 발명에 의하면 타겟 공간별로 발생하는 사운드를 수집하여 분석하므로 타겟 공간의 사운드 환경에 특화된 사운드 분석이 가능하다.
- [0017] 또한, 타겟 공간의 사운드 환경에 특화된 사운드 학습 모델을 이용하여 타겟 공간에서 발생하는 사운드를 효과적으로 학습할 수 있다.
- [0018] 또한, 타겟 공간에서 반복적으로 발생하는 사운드 패턴을 추출하고 이를 이용하여 사운드 학습 모델을 훈련하므로 해당 타겟 공간에서 수집한 입력 사운드로부터 효과적인 노이즈 제거가 가능하다.
- [0019] 본 발명의 효과는 이상에서 언급된 것들에 한정되지 않으며, 언급되지 아니한 다른 효과들은 아래의 기재로부터 본 발명이 속한 기술 분야의 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0020] 도 1은 일 실시예에 따른 사운드 분석 환경의 예시도이다.
- 도 2는 일 실시예에 따른 사운드 분석 장치의 개략적인 블록도이다.
- 도 3은 일 실시예에 따른 사운드 분석 방법의 흐름도이다.
- 도 4 및 도 5는 일 실시예에 따른 사운드 분할 방식을 예시적으로 설명하기 위한 도면이다.
- 도 6은 일 실시예에 따른 반복 매칭 관계를 설명하기 위한 도면이다.
- 도 7은 일 실시예에 따른 반복 사운드 패턴을 설명하기 위한 도면이다.
- 도 8은 일 실시예에 따른 타겟 사운드 분석 과정의 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0021] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시예를 상세히 설명하되, 동일하거나 유사한 구성요소에는 동일유사한 도면 부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 또한, 본 명세서에 개시된 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 실시예의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0022] 본 출원에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다. 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0023] 실시예를 설명하기에 앞서 본 발명의 배경에 대하여 우선 살펴보기로 한다.
- [0024] 소리 정보는 동시 다발적으로 발생하게 되기 때문에 목적에 따라 노이즈가 데이터일 수 있고 데이터가 노이즈일 수 있다. 구체적으로, 대화 내용을 분석할 때에는 사람의 음성이 중요 데이터이고 그 외의 소리가 노이즈로 처리되지만, 반대로 상황 정보를 분석할 때에는 그 외의 소리들이 중요한 데이터이고 사람의 음성이 노이즈로 처리될 것이다. 그러나, 공간에 따라서 노이즈가 발생하는 환경과 조건 혹은 노이즈의 종류가 다르기 때문에 단일 모델을 통해 분석하기에 어려움이 있다. 이는 동일한 내용의 대화라도 주변 상황 혹은 노이즈 제거(noise reduction) 기법에 따라서 노이즈로 필터링될 수도 있고 유의미한 데이터가 될 수도 있기 때문이다.
- [0025] 따라서, 오디오 데이터를 분석하기 전에 노이즈/데이터를 올바르게 분별하기 위한 전처리 기술이 필요하다.
- [0026] 일반적으로 오디오는 어떤 공간인지에 따라서 소리의 종류와 크기, 주로 발생하는 소리 등 여러 특성이 달라지게 된다. 그렇기 때문에 마이크가 고정된 장소에 존재하게 되면 노이즈/데이터를 올바르게 판별하기 어려울 수 있다. 이를 위해, 하나의 마이크가 아니라 공간마다 별도의 마이크를 두어 해당 공간에서 발생하는 소리에 대해서 효과적으로 학습하고, 해당 공간의 반복적인 소리 패턴 등을 고려하여 노이즈를 정확하게 제거할 수 있는 사운드 분석 방법 및 장치가 필요하다. 실시예에 따른 사운드 분석 방법 및 장치는 타겟 장소의 사운드 환경에 따라 노이즈/데이터를 올바르게 분별할 수 있는 전처리를 제공한다.
- [0027] 도 1은 일 실시예에 따른 사운드 분석 환경의 예시도이다.
- [0028] 사운드 분석 환경은, 사운드 분석 장치(100), 서버(200) 및 네트워크(300)를 포함할 수 있다. 사운드 분석 장치(100) 및 서버(200)는 5G 통신 환경에서 서로 연결될 수 있다. 또한, 도 1에 도시된 기기들 이외에 가정 또는 사무실에서 사용되는 다양한 전자 기기들이 사물 인터넷 환경 하에서 서로 연결되어 동작할 수 있다.
- [0029] 사운드 분석 장치(100)는 다양한 타겟 공간에서 입력 사운드를 획득하고 획득한 사운드를 분석한다. 예를 들어, 사운드 분석 장치(100)는 인공지능(AI) 스피커 또는 커뮤니케이션 로봇 등과 같은 음성 인식 기능을 수행할 수 있는 다양한 전자 기기들을 포함할 수 있다. 예를 들어, 사운드 분석 장치(100)는 데스크 탑 컴퓨터, 스마트폰, 노트북, 태블릿 PC, 스마트 TV, 휴대폰, PDA(personal digital assistant), 랩톱, 미디어 플레이어, 마이크로 서버, GPS(global positioning system) 장치, 전자책 단말기, 디지털 방송용 단말기, 네비게이션, 키오스크, MP3 플레이어, 디지털 카메라, 가전기기 및 기타 모바일 또는 비모바일 컴퓨팅 장치일 수 있으나, 이에 제한되지 않는다. 또한, 사운드 분석 장치(100)는 통신 기능 및 데이터 프로세싱 기능을 구비한 시계, 안경, 헤어 밴

드 및 반지 등의 웨어러블 디바이스일 수 있다. 예를 들어, 사운드 분석 장치(100)는 TV, 세탁기, 에어컨 등과 같은 음성 인식 기능을 수행할 수 있는 다양한 가전 기기들을 포함할 수 있다. 또한 사운드 분석 장치(100)는 음성 입출력 기능이 없는 전자 기기를 제어하는 허브 역할을 할 수 있다.

- [0030] 사운드 분석 장치(100)는 자체적으로 사운드 분석 기능을 수행할 수 있고, 또는 서버(200)에 입력 사운드 또는 타겟 사운드를 전송하고 서버(200)로부터 사운드 분석 결과를 수신할 수 있다. 사운드 분석 장치(100)는 사운드 분석의 결과 결정된 반복 사운드 패턴 정보를 활용하여 음성 인식 서비스의 정확도를 높일 수 있다.
- [0031] 서버(200)는 각종 인공 지능 알고리즘을 적용하는 데에 필요한 빅데이터 및 사운드 인식 및 분석에 관한 데이터를 제공하는 데이터베이스 서버일 수 있다. 그 밖에 서버(200)는 사운드 분석 장치(100)를 원격에서 제어할 수 있도록 하는 웹 서버 또는 애플리케이션 서버를 포함할 수 있다.
- [0032] 인공 지능은 인공적인 지능 또는 이를 만들 수 있는 방법론을 연구하는 분야를 의미하며, 머신 러닝(기계 학습, Machine Learning)은 인공 지능 분야에서 다루는 다양한 문제를 정의하고 그것을 해결하는 방법론을 연구하는 분야를 의미한다. 머신 러닝은 어떠한 작업에 대하여 꾸준한 경험을 통해 그 작업에 대한 성능을 높이는 알고리즘으로 정의하기도 한다.
- [0033] 인공 신경망(ANN: Artificial Neural Network)은 머신 러닝에서 사용되는 모델로서, 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)들로 구성되는, 문제 해결 능력을 가지는 모델 전반을 의미할 수 있다. 인공 신경망은 다른 레이어의 뉴런들 사이의 연결 패턴, 모델 파라미터를 갱신하는 학습 과정, 출력값을 생성하는 활성화 함수(Activation Function)에 의해 정의될 수 있다.
- [0034] 인공 신경망은 입력층(Input Layer), 출력층(Output Layer), 그리고 선택적으로 하나 이상의 은닉층(Hidden Layer)를 포함할 수 있다. 각 층은 하나 이상의 뉴런을 포함하고, 인공 신경망은 뉴런과 뉴런을 연결하는 시냅스를 포함할 수 있다. 인공 신경망에서 각 뉴런은 시냅스를 통해 입력되는 입력 신호들, 가중치, 편향에 대한 활성화 함수의 함수값을 출력할 수 있다.
- [0035] 모델 파라미터는 학습을 통해 결정되는 파라미터를 의미하며, 시냅스 연결의 가중치와 뉴런의 편향 등이 포함된다. 그리고, 하이퍼파라미터는 머신 러닝 알고리즘에서 학습 전에 설정되어야 하는 파라미터를 의미하며, 학습률(Learning Rate), 반복 횟수, 미니 배치 크기, 초기화 함수 등이 포함된다.
- [0036] 인공 신경망의 학습의 목적은 손실 함수를 최소화하는 모델 파라미터를 결정하는 것으로 볼 수 있다. 손실 함수는 인공 신경망의 학습 과정에서 최적의 모델 파라미터를 결정하기 위한 지표로 이용될 수 있다.
- [0037] 머신 러닝은 학습 방식에 따라 지도 학습(Supervised Learning), 비지도 학습(Unsupervised Learning), 강화 학습(Reinforcement Learning)으로 분류할 수 있다.
- [0038] 지도 학습은 학습 데이터에 대한 레이블(label)이 주어진 상태에서 인공 신경망을 학습시키는 방법을 의미하며, 레이블이란 학습 데이터가 인공 신경망에 입력되는 경우 인공 신경망이 추론해 내야 하는 정답(또는 결과 값)을 의미할 수 있다. 비지도 학습은 학습 데이터에 대한 레이블이 주어지지 않는 상태에서 인공 신경망을 학습시키는 방법을 의미할 수 있다. 강화 학습은 어떤 환경 안에서 정의된 에이전트가 각 상태에서 누적 보상을 최대화하는 행동 혹은 행동 순서를 선택하도록 학습시키는 학습 방법을 의미할 수 있다.
- [0039] 인공 신경망 중에서 복수의 은닉층을 포함하는 심층 신경망(DNN: Deep Neural Network)으로 구현되는 머신 러닝을 딥 러닝(심층 학습, Deep Learning)이라 부르기도 하며, 딥 러닝은 머신 러닝의 일부이다. 이하에서, 머신 러닝은 딥 러닝을 포함하는 의미로 사용된다.
- [0040] 서버(200)는 사운드 분석 장치(100)로부터 수신한 사운드를 분석하고, 사운드 분석 결과를 사운드 분석 장치(100)로 전송할 수 있다. 서버(200)는 사운드 분석 장치(100)로부터 분석 대상이 되는 사운드 정보를 수신하고, 수신된 사운드의 반복 사운드 패턴을 결정할 수 있다. 이를 위하여 서버(200)는 머신 러닝에 기반한 사운드 학습 모델을 이용할 수 있다. 서버(200)는 결정된 반복 사운드 패턴을 사운드 분석 장치(100)로 전송할 수 있다. 서버(200)는 결정된 반복 사운드 패턴을 사운드 학습 모델에 입력 데이터로서 제공하여 사운드 학습 모델을 훈련 및 갱신할 수 있다.
- [0041] 서버(200)는 타겟 공간의 입력 사운드 또는 타겟 사운드를 사운드 분석 장치(100)로부터 수신하고 수신된 사운드를 타겟 공간 정보와 함께 서버(200)가 접근가능한 저장소에 저장 및 관리할 수 있다. 또한 서버(200)는 사운드 분석 결과 결정된 타겟 공간의 반복 사운드 패턴을 타겟 공간 정보와 함께 저장소에 저장 및 관리할 수

있다.

- [0042] 네트워크(300)는 사운드 분석 장치(100)와 서버(200)를 연결하는 역할을 수행할 수 있다. 이러한 네트워크(300)는 예컨대 LANs(local area networks), WANs(wide area networks), MANs(metropolitan area networks), ISDNs(integrated service digital networks) 등의 유선 네트워크나, 무선 LANs, CDMA, 블루투스, 위성 통신 등의 무선 네트워크를 망라할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다. 또한 네트워크(300)는 근거리 통신 및/또는 원거리 통신을 이용하여 정보를 송수신할 수 있다. 여기서 근거리 통신은 블루투스(Bluetooth), RFID(radio frequency identification), 적외선 통신(IrDA, infrared data association), UWB(ultra-wideband), ZigBee, Wi-Fi (wireless fidelity) 기술을 포함할 수 있고, 원거리 통신은 CDMA(code division multiple access), FDMA(frequency division multiple access), TDMA(time division multiple access), OFDMA(orthogonal frequency division multiple access), SC-FDMA(single carrier frequency division multiple access) 기술을 포함할 수 있다.
- [0043] 네트워크(300)는 허브, 브리지, 라우터, 스위치 및 게이트웨이와 같은 네트워크 요소들의 연결을 포함할 수 있다. 네트워크(300)는 인터넷과 같은 공용 네트워크 및 안전한 기업 사설 네트워크와 같은 사설 네트워크를 비롯한 하나 이상의 연결된 네트워크들, 예컨대 다중 네트워크 환경을 포함할 수 있다. 네트워크(300)에의 액세스는 하나 이상의 유선 또는 무선 액세스 네트워크들을 통해 제공될 수 있다. 더 나아가 네트워크(300)는 사물 등 분산된 구성 요소들 간에 정보를 주고 받아 처리하는 IoT(Internet of Things, 사물인터넷) 망 및/또는 5G 통신을 지원할 수 있다.
- [0044] 도 2는 일 실시예에 따른 사운드 분석 장치의 개략적인 블록도이다.
- [0045] 사운드 분석 장치(100)는 통신부(110), 디스플레이부(121) 및 조작부(122)를 포함하는 사용자 인터페이스부(120), 센싱부(130), 오디오 입력부(141) 및 오디오 출력부(142)를 포함하는 오디오 처리부(140), 메모리(150) 및 프로세서(160)를 포함할 수 있다.
- [0046] 통신부(110)는 네트워크(300)와 연동하여 사운드 분석 장치(100)와 서버(200) 간의 송수신 신호를 패킷 데이터 형태로 제공하는 데에 필요한 통신 인터페이스를 제공할 수 있다. 또한 통신부(110)는 다른 네트워크 장치와 유무선 연결을 통해 제어 신호 또는 데이터 신호와 같은 신호를 송수신하기 위해 필요한 하드웨어 및 소프트웨어를 포함하는 장치일 수 있다. 또한, 통신부(110)는 각종 사물 지능 통신(IoT(internet of things), IoE(internet of everything), IoST(internet of small things) 등)을 지원할 수 있으며, M2M(machine to machine) 통신, V2X(vehicle to everything communication) 통신, D2D(device to device) 통신 등을 지원할 수 있다.
- [0047] 통신부(110)는 프로세서(160)의 제어 하에 입력 사운드를 포함하는 데이터를 수신할 수 있다. 프로세서(160)는 통신부(110)를 통해 수신한 입력 사운드를 포함하는 데이터로부터 입력 사운드를 추출하여 실시예에 따른 사운드 분석 방법을 수행할 수 있다.
- [0048] 통신부(110)는 프로세서(160)의 제어 하에 서버(200)로 입력 사운드를 전송하고, 이에 대한 응답으로 서버(200)로부터 사운드 분석 결과를 수신할 수 있다. 여기서 사운드 분석 결과는 예를 들어 입력 사운드의 반복 매칭 관계, 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴 및 사운드 학습 모델을 포함할 수 있다.
- [0049] 사용자 인터페이스부(120) 중 디스플레이부(121)는 프로세서(160)의 제어 하에 사운드 분석 장치(100)의 구동 상태를 디스플레이 할 수 있다. 일 예에서 디스플레이부(121)는 터치패드와 상호 레이어 구조를 이루어 터치스크린으로 구성될 수 있다. 이 경우에, 디스플레이부(121)는 사용자의 터치에 의한 정보의 입력이 가능한 조작부(122)로도 사용될 수 있다. 이를 위해 디스플레이부(121)는 터치 인식 디스플레이 제어기 또는 이외의 다양한 입출력 제어기로 구성될 수 있다. 일 예로, 터치 인식 디스플레이 제어기는 장치와 사용자 사이에 출력 인터페이스 및 입력 인터페이스를 제공할 수 있다. 터치 인식 디스플레이 제어기는 전기 신호를 프로세서(160)와 송수신할 수 있다. 또한, 터치 인식 디스플레이 제어기는 사용자에게 시각적인 출력을 표시하며, 시각적 출력은 텍스트, 그래픽, 이미지, 비디오와 이들의 조합을 포함할 수 있다. 이와 같은 디스플레이부(121)는 예를 들어 터치 인식이 가능한 OLED(organic light emitting display) 또는 LCD(liquid crystal display) 또는 LED(light emitting display)와 같은 소정의 디스플레이 부재일 수 있다.
- [0050] 사용자 인터페이스부(120) 중 조작부(122)는 복수의 조작 버튼(미도시)을 구비하여, 입력되는 버튼에 대응하는 신호를 프로세서(160)로 전송할 수 있다. 이러한 조작부(122)는 사용자의 터치 또는 누름 조작을 인식할 수 있는 센서 또는 버튼, 스위치 구조로 구성될 수 있다. 조작부(122)는 디스플레이부(121)에 표시되는 사운드 분석

장치(100)의 구동과 관련한 각종 정보를 확인하거나, 변경하기 위해 사용자가 조작하는 조작 신호를 프로세서(160)로 전송할 수 있다.

- [0051] 디스플레이부(121)는 프로세서(160)의 제어 하에 사운드 분석 장치(100)의 사운드 분석 결과를 출력할 수 있다. 여기서 사운드 분석 결과는 예를 들어 입력 사운드의 반복 매칭 관계, 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴 및 사운드 학습 모델을 포함할 수 있다.
- [0052] 센싱부(130)는 사운드 분석 장치(100)의 주변 상황을 센싱하는 각종 센서를 포함할 수 있다.
- [0053] 센싱부(130)는 위치 센서를 포함할 수 있다. 위치 센서는 사운드 분석 장치(100)의 위치 데이터를 획득할 수 있다. 예를 들어 위치 센서는 GPS(Global Positioning System) 센서 및 IMU(Inertial Measurement Unit) 센서 등을 포함한다. IMU 센서는 예를 들어, 가속도 센서, 및/또는 자이로 센서 등을 포함할 수 있다. 위치 센서가 획득한 위치 데이터는 메모리(150)에 저장될 수 있다.
- [0054] 센싱부(130)는 근접 센서를 포함할 수 있다. 근접 센서는 적외선 등을 활용하여 사운드 분석 장치(100) 주변에 위치한 객체(예를 들어 사용자)의 위치 데이터를 획득할 수 있다. 한편, 근접 센서가 획득한 사용자의 위치 데이터는 메모리(150)에 저장될 수 있다.
- [0055] 센싱부(130)는 영상 센서를 포함할 수 있다. 영상 센서는 사운드 분석 장치(100) 주변을 촬영할 수 있는 카메라(미도시)를 포함할 수 있으며, 촬영 효율을 위해 복수 개가 설치될 수 있다. 예를 들어, 카메라는 적어도 하나의 광학렌즈와, 광학렌즈를 통과한 광에 의해 상이 맺히는 다수개의 광다이오드(photodiode, 예를 들어, pixel)를 포함하여 구성된 이미지 센서(예를 들어, CMOS image sensor)와, 광다이오드들로부터 출력된 신호를 바탕으로 영상을 구성하는 디지털 신호 처리기(DSP: digital signal processor)를 포함할 수 있다. 디지털 신호 처리기는 정지영상은 물론이고, 정지영상으로 구성된 프레임들로 이루어진 동영상상을 생성할 수 있다. 한편, 영상 센서로서의 카메라가 촬영하여 획득된 영상은 메모리(150)에 저장될 수 있다.
- [0056] 센싱부(130)는 사운드 분석 장치(100)의 주변 상황을 감지할 수 있는 센서, 예를 들어, 라이다 센서(Lidar sensor), 무게 감지 센서, 조도 센서(illumination sensor), 터치 센서(touch sensor), 가속도 센서(acceleration sensor), 자기 센서(magnetic sensor), 중력 센서(G-sensor), 자이로스코프 센서(gyroscope sensor), 모션 센서(motion sensor), RGB 센서, 적외선 센서(IR 센서: infrared sensor), 지문인식 센서(finger scan sensor), 초음파 센서(ultrasonic sensor), 광 센서(optical sensor), 마이크로폰(microphone), 배터리 게이지(battery gauge), 환경 센서(예를 들어, 기압계, 습도계, 온도계, 방사능 감지 센서, 열 감지 센서, 가스 감지 센서 등), 화학 센서(예를 들어, 전자 코, 헬스케어 센서, 생체 인식 센서 등) 중 적어도 하나를 포함할 수 있다. 한편, 사운드 분석 장치(100)는 이러한 센서들 중 적어도 둘 이상의 센서에서 센싱되는 정보를 조합하여 활용할 수 있다.
- [0057] 오디오 처리부(140) 중 오디오 입력부(141)는 타겟 공간에서 발생하는 입력 사운드를 수신할 수 있다. 이를 위해 오디오 입력부(141)는 하나 이상의 마이크로폰(미도시)을 포함할 수 있다. 오디오 입력부(141)는 마이크로폰으로 감지한 사운드를 프로세서(160)에 전달할 수 있다.
- [0058] 오디오 입력부(141)는 복수의 마이크로폰(미도시)을 구비할 수 있다. 여기서, 복수의 마이크로폰 각각은 서로 다른 타겟 공간에 이격되어 배치될 수 있고, 수신한 사운드를 전기적인 신호로 처리할 수 있다. 오디오 입력부(141)는 서로 다른 타겟 공간에 배치된 복수의 마이크로폰으로부터 입력 사운드를 획득할 수 있다. 복수의 마이크로폰을 이용하여 각 타겟 공간에서 수집된 복수의 입력 사운드는 각 입력 사운드가 수신된 타겟 공간 정보와 함께 메모리(150)에 저장된다.
- [0059] 일 예에서 타겟 공간마다 마이크로폰을 배치하는 대신에 사운드 분석 장치(100)가 각 타겟 공간을 방문하여 입력 사운드를 획득할 수 있다. 예를 들어 사운드 분석 장치(100)는 주기적으로 각 타겟 공간에 소정의 기간 동안 머물면서 입력 사운드를 획득할 수 있다. 이 경우, 사운드 분석 장치(100)는 이동이 가능한 전자 장치로서, 예를 들어 이동가능한 홈 로봇일 수 있다.
- [0060] 선택적으로 오디오 입력부(141)는 입력 사운드를 수신하는 과정에서 발생하는 노이즈를 제거하기 위한 다양한 노이즈 제거 알고리즘을 사용할 수 있다. 선택적 실시 예로 오디오 입력부(141)는 입력 사운드 수신 시에 노이즈를 제거하는 필터(미도시), 필터에서 출력되는 신호를 증폭하여 출력하는 증폭기(미도시) 등 음성 신호 처리를 위한 각종 구성 요소들을 포함할 수 있다.
- [0061] 오디오 처리부(140) 중 오디오 출력부(142)는 프로세서(160)의 제어에 따라 경고음, 동작모드, 동작상태, 예러

상태 등의 알림 메시지, 입력 사운드에 대한 분석 결과, 사용자의 발화에 대응하는 응답 정보, 사용자의 음성 입력에 대응하는 처리 결과 등을 오디오로 출력할 수 있다. 오디오 출력부(142)는 프로세서(160)로부터의 전기 신호를 오디오 신호로 변환하여 출력할 수 있다. 이를 위해, 오디오 출력부(142)는 스피커를 포함한다. 오디오 출력부(142)는 프로세서(160)의 제어 하에 스피커를 통해 사운드 분석 장치(100)의 분석 결과를 출력할 수 있다. 여기서 분석 결과는 예를 들어 입력 사운드의 반복 매칭 관계 및 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 포함할 수 있다.

[0062] 메모리(150)는 사운드 분석 장치(100)의 동작에 필요한 각종 정보들을 저장하는 것으로, 휘발성 또는 비휘발성 기록 매체를 포함할 수 있다. 예를 들어, 메모리(150)는 사운드 분석 장치(100)를 활성화하는 기동어를 포함한 음성 명령어를 저장할 수 있다.

[0063] 메모리(150)는 프로세서(160)의 제어 하에 오디오 입력부(141)를 통하여 수신되는 입력 사운드를 저장할 수 있다. 또한 메모리(150)는 센싱부(130)가 감지한 정보를 저장할 수 있다. 메모리(150)는 프로세서(160)의 제어 하에 사운드 분석 장치(100)가 분석한 각종 정보를 저장할 수 있다. 사운드 분석 장치(100)가 분석한 정보는 예를 들어 입력 사운드의 반복 매칭 관계, 타겟 공간의 반복 사운드 패턴 정보 및 사운드 학습 모델을 포함할 수 있다.

[0064] 메모리(150)는 프로세서(160)의 제어 하에 사운드 학습 모델을 저장할 수 있다.

[0065] 메모리(150)는 자기 저장 매체(magnetic storage media) 또는 플래시 저장 매체(flash storage media)를 포함할 수 있으나, 이에 한정되는 것은 아니다. 이러한 메모리(150)는 내장 메모리 및/또는 외장 메모리를 포함할 수 있으며, DRAM, SRAM, 또는 SDRAM 등과 같은 휘발성 메모리, OTPROM(one time programmable ROM), PROM, EPROM, EEPROM, mask ROM, flash ROM, NAND 플래시 메모리, 또는 NOR 플래시 메모리 등과 같은 비휘발성 메모리, SSD, CF(compact flash) 카드, SD 카드, Micro-SD 카드, Mini-SD 카드, Xd 카드, 또는 메모리 스틱(memory stick) 등과 같은 플래시 드라이브, 또는 HDD와 같은 저장 장치를 포함할 수 있다.

[0066] 프로세서(160)는 오디오 입력부(141)를 통하여 수신한 입력 사운드를 분석하고 그 결과를 디스플레이부(121)를 통하여 시각적 정보로 제공하거나, 오디오 출력부(142)를 통하여 청각적 정보로 제공할 수 있다.

[0067] 프로세서(160)는 일종의 중앙처리장치로서 메모리(150)에 탑재된 제어 소프트웨어를 구동하여 사운드 처리 장치(100)의 전체 동작을 제어할 수 있다. 프로세서(160)는 데이터를 처리할 수 있는 모든 종류의 장치를 포함할 수 있다. 프로세서(160)는 예를 들어 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 이와 같이 하드웨어에 내장된 데이터 처리 장치의 일 예로써, 마이크로프로세서(microprocessor), 중앙처리장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티프로세서(multiprocessor), ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 처리 장치를 망라할 수 있으나, 이에 한정되는 것은 아니다.

[0068] 프로세서(160)는 사운드 분석 장치(100)의 각 구성 모듈을 제어하여 실시예에 따른 사운드 분석 방법을 구현한 응용 프로그램을 실행할 수 있다.

[0069] 프로세서(160)는 오디오 입력부(141)의 마이크로폰을 제어하여 타겟 공간의 입력 사운드를 획득하거나 통신부(110)를 제어하여 원격에서 획득된 입력 사운드를 수신할 수 있다.

[0070] 프로세서(160)는 사운드 학습 모델에 기반하여 타겟 공간에서 획득한 타겟 사운드 내의 반복 사운드 패턴을 결정하도록 동작할 수 있다. 프로세서(160)는 메모리(150)에 저장된 사운드 학습 모델을 이용하여 타겟 사운드 내의 반복 사운드 패턴을 결정하도록 동작할 수 있다. 프로세서(160)는 통신부(110)를 통해 타겟 사운드를 서버(200)에 전송하고 서버(200)에 저장된 사운드 학습 모델을 이용하여 서버(200)가 결정한 타겟 사운드의 반복 사운드 패턴을 통신부(110)를 통해 수신할 수 있다.

[0071] 프로세서(160)는 타겟 공간에서 발생한 입력 사운드를 획득하고, 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하도록 설정될 수 있다. 프로세서(160)는 입력 사운드의 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하도록 설정될 수 있다. 프로세서(160)는 반복 매칭 관계에 기반하여 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하도록 설정될 수 있다.

[0072] 프로세서(160)는 입력 사운드에서 추출된 반복 사운드 패턴에 기반하여 메모리(150)에 저장된 사운드 학습 모델을 훈련시키도록 설정될 수 있다. 일 예에서, 프로세서(160)는 입력 사운드에서 추출된 반복 사운드 패턴을 통

신부(110)를 통해 서버(200)로 전송할 수 있다.

- [0073] 프로세서(160)는 서버(200)에서 훈련된 사운드 학습 모델을 통신부(110)를 통해 수신하여 메모리(150)에 저장할 수 있다.
- [0074] 프로세서(160)는 타겟 공간에서 타겟 사운드를 획득하고, 타겟 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 타겟 서브 사운드를 생성하도록 설정될 수 있다. 프로세서(160)는 메모리(150)에 저장된 사운드 학습 모델에 기반하여 각 타겟 서브 사운드가 타겟 공간의 반복 사운드 패턴인 지 여부를 결정하도록 설정될 수 있다.
- [0075] 프로세서(160)는 타겟 공간의 반복 사운드 패턴인 것으로 결정된 타겟 서브 사운드를 타겟 공간의 노이즈로 결정하도록 설정될 수 있다. 프로세서(160)는 타겟 공간의 노이즈로 결정된 타겟 서브 사운드를 음성 인식 동작에 활용할 수 있다. 이로써 타겟 공간의 사운드 환경을 고려한 음성 인식 서비스의 제공이 가능하고, 음성 인식의 정확도가 제고될 수 있다.
- [0076] 프로세서(160)는 사운드 분석 장치(100)가 최적의 분석 결과를 출력하도록, 획득한 사운드에 대하여 딥러닝(Deep Learning) 등 머신 러닝(machine learning)을 수행할 수 있고, 메모리(150)는, 머신 러닝에 사용되는 데이터, 결과 데이터 등을 저장할 수 있다. 예를 들어 메모리(150)는 획득한 사운드에서 반복 사운드 패턴을 판별하기 위한 사운드 학습 모델을 저장할 수 있다.
- [0077] 프로세서(160)는 인공신경망(artificial neural network)을 구동할 수 있고, 사운드 데이터를 입력 데이터로 하는 머신 러닝 기반의 사운드 분석을 수행할 수 있다. 예를 들어 프로세서(160)는 머신 러닝에 기반한 사운드 학습 모델을 이용하여 타겟 공간에서 발생하는 사운드를 분석하여 사운드 내의 반복 매칭 관계를 결정하고 반복 사운드 패턴을 판별할 수 있다.
- [0078] 도 3은 일 실시예에 따른 사운드 분석 방법의 흐름도이다.
- [0079] 사운드 분석 장치(100)는 실시예에 따른 사운드 분석 방법에 따라 타겟 공간에서 발생하는 사운드를 수집, 분석 및 학습할 수 있다.
- [0080] 사운드 분석 방법은 타겟 공간에서 발생한 입력 사운드를 획득하는 단계(310), 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성하는 단계(320), 서브 사운드 간의 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정하는 단계(330), 반복 매칭 관계에 기반하여 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출하는 단계(340) 및 반복 사운드 패턴에 기반하여 타겟 공간에 대한 사운드 학습 모델을 훈련하는 단계(350)를 포함할 수 있다.
- [0081] 단계(310)에서 사운드 분석 장치(100)는 타겟 공간에서 발생한 입력 사운드를 획득할 수 있다. 사운드 분석 장치(100)는 오디오 입력부(141)를 통해 입력 사운드를 직접 획득 및/또는 통신부(110)를 통해 입력 사운드를 포함한 데이터를 수신할 수 있다.
- [0082] 단계(310)에서 사운드 분석 장치(100)는 타겟 공간에 배치된 마이크로폰을 이용하여 입력 사운드를 감지할 수 있다.
- [0083] 타겟 공간은 사운드 분석의 대상이 되는 사운드가 수집된 장소를 의미한다. 일 예에서, 타겟 공간은 일정 면적을 가진 분리된 공간으로서, 고유의 사운드 환경을 가질 수 있다. 예를 들어, 건물의 한 층에 있는 여러 개의 사무실은 각각 하나의 타겟 공간을 구성할 수 있다. 예를 들어, 가정집의 경우 안방, 거실, 부엌, 침실, 베란다 및 서재는 각각 별개의 타겟 공간이 될 수 있다. 일 예에서, 타겟 공간은 하나의 공간적 위치를 의미할 수 있다. 예를 들어, 침실의 경우 창가와 문가는 다른 타겟 공간이 될 수 있다.
- [0084] 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 센싱부(130)를 통해 감지된 타겟 공간의 위치 정보 및/또는 영상 정보를 타겟 공간에 대한 식별 정보로 사용할 수 있다. 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 오디오 입력부(141)의 마이크로폰으로 감지한 사용자의 발화로부터 타겟 공간 식별 정보를 획득하거나 통신부(110)를 통해 타겟 공간 식별 정보를 획득할 수 있다.
- [0085] 사운드 분석 장치(100)는 타겟 공간마다 배치된 복수의 마이크로폰으로부터 입력 사운드를 감지할 수 있다. 사운드 분석 장치(100)는 복수의 타겟 공간에 각각 배치된 마이크로폰을 이용하여 복수의 타겟 공간에서 발생하는 입력 사운드를 공간별로 분석할 수 있다. 일 예에서 사운드 분석 장치(100)는 복수의 타겟 공간을 방문하여 입력 사운드를 감지할 수 있다. 일 예에서 사운드 분석 장치(100)는 타겟 공간마다 배치될 수 있다.
- [0086] 단계(310)에서 사운드 분석 장치(100)는 타겟 공간에서 소정의 시간 주기 마다 입력 사운드를 획득할 수 있다.

예를 들어 소정의 시간 주기는 12시간 또는 24시간일 수 있고, 이에 제한되지 않으며 사용자가 필요에 따라 조정할 수 있다. 사운드 분석 장치(100)는 타겟 공간에서 소정의 시간 주기 동안에 입력 사운드를 반복적으로 획득할 수 있다. 예를 들어, 사운드 분석 장치(100)는 타겟 공간에 배치된 마이크로폰을 통해 타겟 공간에서 발생하는 사운드를 12시간 동안 또는 24시간동안 수집할 수 있다.

- [0087] 단계(320)에서 사운드 분석 장치(100)는 단계(310)에서 획득한 입력 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 서브 사운드를 생성할 수 있다.
- [0088] 입력 사운드를 분할한다는 것은 입력 사운드를 소정의 시간 간격에 따라 복수의 구간으로 구분한다는 것을 의미한다. 입력 사운드의 서브 사운드는 입력 사운드의 일부로서, 입력 사운드의 분할에 사용된 시간 간격 이하의 재생 시간을 갖는다.
- [0089] 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 입력 사운드 파일에서 서브 사운드의 시작 위치 정보와 종료 위치 정보의 쌍(pair)을 메모리(150)에 저장할 수 있다. 이 경우 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 일 세트의 서브 사운드의 각 사운드의 시작 위치 정보와 종료 위치 정보의 쌍을 테이블로 메모리(150)에 저장 및 관리할 수 있다. 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 이와 같은 테이블을 이용하여 각 사운드의 시작 위치 정보와 종료 위치 정보의 쌍에 접근할 수 있다. 여기서 테이블은 사전정의된 인덱스를 이용하여 인덱스에 매핑된 데이터에 접근가능한 데이터 구조로서, 예를 들어 배열(array), 연결 리스트(linked list) 및 해시 테이블(hash table) 등으로 구현될 수 있으며, 이에 제한되지 않는다.
- [0090] 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 각 서브 사운드를 하나의 오디오 파일로 메모리(150)에 저장할 수 있다. 여기서 mp3, wav, wma, avi 및 aac 등과 같은 다양한 오디오 파일 포맷이 사용될 수 있다.
- [0091] 단계(320)에서 사운드 분석 장치(100)는 입력 사운드를 분할하여 생성한 일 세트의 서브 사운드의 각 서브 사운드를 또다른 소정의 시간 간격에 따라 반복적으로 분할하여 또다른 세트의 서브 사운드를 생성할 수 있다. 또한, 단계(320)에서 사운드 분석 장치(100)는 일 세트의 서브 사운드를 생성한 분할과는 상이한 시작 지점부터 입력 사운드를 소정의 시간 간격에 따라 분할하여 또다른 세트의 서브 사운드를 생성할 수 있다.
- [0092] 말하자면, 단계(320)에서 사운드 분석 장치(100)는 복수의 분할 방식을 이용하여 입력 사운드를 반복적으로 분할하여 복수의 서브 사운드 세트를 생성할 수 있다.
- [0093] 복수의 분할 방식은 분할 시간 간격을 달리하거나 분할의 시작 지점을 달리하는 분할 방식들을 포함한다. 이하에서 도 4 및 도 5를 참조하여 예시적인 사운드 분할 방식을 살펴본다.
- [0094] 도 4 및 도 5는 일 실시예에 따른 사운드 분할 방식을 예시적으로 설명하기 위한 도면이다.
- [0095] 사운드 분석 장치(100)는 서브 사운드의 시간 길이가 임계 길이 이하가 될 때까지 입력 사운드를 반복적으로 N등분할 수 있다(여기서 N은 1보다 큰 자연수). 임계 길이는 시간 단위로 설정되고, 예를 들어 1시간일 수 있으며, 사용자가 필요에 따라 조정가능하다.
- [0096] 도 4는 입력 사운드(S₁)를 반복적으로 2등분하여 복수의 서브 사운드 세트를 생성하는 분할 방식을 예시한다.
- [0097] 제 1 세트의 서브 사운드는 입력 사운드(S₁)을 2등분하여 생성된 서브 사운드(S₂) 및 서브 사운드(S₃)를 포함한다.
- [0098] 제 2 세트의 서브 사운드는 서브 사운드(S₂) 및 서브 사운드(S₃)를 각각 다시 2등분하여 생성된 서브 사운드(S₄), 서브 사운드(S₅), 서브 사운드(S₆) 및 서브 사운드(S₇)를 포함한다.
- [0099] 제 3 세트의 서브 사운드는 제 2 세트의 서브 사운드에 포함된 각 서브 사운드(S₄, S₅, S₆, S₇)를 각각 다시 2등분하여 생성된 서브 사운드(S₈, S₉, S₁₀, S₁₁, S₁₂, S₁₃, S₁₄, S₁₅)를 포함한다.
- [0100] 사운드 분석 장치(100)는 제 3 세트의 각 서브 사운드의 시간 길이가 사전설정된 임계 길이보다 긴 경우, 추가인 분할을 반복하여 추가 세트의 서브 사운드를 생성할 수 있다.
- [0101] 도 5(a)는 입력 사운드(S₁)를 반복적으로 3등분하여 일 세트의 서브 사운드 세트(S₂, S₃, S₄)를 생성하는 분할 방식을 예시한다.
- [0102] 도 5(b)는 입력 사운드의 분할 시작 지점을 달리하여 일 세트의 서브 사운드를 생성하는 방식을 예시한다.

- [0103] 사운드 분석 장치(100)는 입력 사운드의 제 1 시작 지점(t_0)부터 소정의 시간 간격에 따라 분할하여 제 1 세트의 서브 사운드를 생성한다. 사운드 분석 장치(100)는 입력 사운드의 제 2 시작 지점(t_0')부터 소정의 시간 간격에 따라 분할하여 제 2 세트의 서브 사운드를 생성한다. 일 예에서 제 2 시작 지점(t_0')은 사전설정된 임계 길이보다 작은 값을 가질 수 있다. 일 예에서 제 2 시작 지점(t_0')은 제 1 시작 지점(t_0)부터 소정의 시간 간격에 따라 분할하여 생성된 제 1 세트의 서브 사운드의 맨 마지막 서브 사운드의 길이로 설정될 수 있다.
- [0104] 사운드 분석 장치(100)는 복수의 분할 방식을 혼용하여 사용할 수 있다. 예를 들어 제 1 세트의 서브 사운드는 입력 사운드를 2등분하고, 제 2 세트의 서브 사운드는 제 1 세트의 서브 사운드를 3등분하고, 제 3 세트의 서브 사운드는 제 2 세트의 서브 사운드를 시작 지점(t_2)을 달리하여 소정의 시간 간격에 따라 분할하고, 제 4 세트의 서브 사운드는 제 2 세트의 서브 사운드를 새로운 시작 지점(t_3)부터 소정의 시간 간격에 따라 분할하여 생성될 수 있다.
- [0105] 단계(330)에서 사운드 분석 장치(100)는 단계(320)에서 생성한 일 세트의 서브 사운드의 서브 사운드 유사도에 기반하여 각 서브 사운드에 대한 반복 매칭 관계를 결정할 수 있다.
- [0106] 사운드 분석 장치(100)는 일 세트의 서브 사운드의 각 서브 사운드에 대하여 반복 매칭 관계를 결정할 수 있다.
- [0107] 이를 위하여 단계(330)는 단계(320)에서 생성한 일 세트의 서브 사운드의 하나의 서브 사운드를 타겟 서브 사운드로 선택하는 단계를 포함할 수 있다. 일 예에서, 사운드 분석 장치(100)는 일 세트의 서브 사운드의 각 서브 사운드를 순차적으로 타겟 서브 사운드로 선택할 수 있다. 일 예에서, 사운드 분석 장치(100)는 타겟 장소에서 수집된 입력 사운드를 분석하여 누적된 로그 정보에 기반하여 일 세트의 서브 사운드 중 반복 매칭 관계가 발생할 가능성이 가장 높은 시간 지점의 서브 사운드를 타겟 서브 사운드로 선택할 수 있다.
- [0108] 반복 매칭 관계는 일 세트의 서브 사운드의 각 서브 사운드(즉, 타겟 서브 사운드)가 일 세트의 서브 사운드 내에서 반복적으로 등장하는 지 여부를 판단하고, 타겟 서브 사운드가 반복적으로 등장하는 경우, 타겟 서브 사운드의 반복 위치를 구조화한 것을 의미한다. 반복 매칭 관계는 시간 길이가 동일한 서브 사운드 간에 정의될 수 있다.
- [0109] 반복 위치는 타겟 서브 사운드가 반복적으로 등장하기 시작하는 시작 위치 정보 및/또는 종료 위치 정보를 포함할 수 있다. 반복 위치는 반복된 서브 사운드가 차지하는 타임 슬롯 정보를 포함할 수 있다. 일 예에서, 타겟 서브 사운드의 반복 매칭 관계는 다양한 데이터 구조로 표현될 수 있다. 예를 들어 반복 매칭 관계는 연결 리스트로 표현될 수 있다.
- [0110] 예를 들어, 1시간 단위로 분할된 일 세트의 서브 사운드에서 제 1 타임 슬롯(0시-1시)에 있는 서브 사운드를 타겟 서브 사운드로 선택한 경우, 타겟 서브 사운드가 제 2 타임 슬롯(3시-4시) 및 제 3 타임 슬롯(12시-13시)에서 반복된다면, 반복 위치는 제 1 타임 슬롯, 제 2 타임 슬롯 및 제 3 타임 슬롯으로 결정될 수 있다. 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 이와 같은 반복 위치를 반복 매칭 관계로 구조화하여 메모리(150)에 연결 리스트로 저장할 수 있다. 여기서 연결 리스트는 제 1 타임 슬롯-제 2 타임 슬롯-제 3 타임 슬롯으로 순차적으로 연결될 수 있다.
- [0111] 서브 사운드 유사도는 서브 사운드들이 서로 매칭되는 정도를 의미한다. 단계(330)에서 사운드 분석 장치(100)는 두 개의 서브 사운드들의 파형을 비교하여 서로 매칭되는 정도에 따라 양 서브 사운드 간의 유사도를 결정할 수 있다.
- [0112] 사운드 분석 장치(100)는 일 세트의 서브 사운드 중 선택한 타겟 서브 사운드와 일 세트의 서브 사운드의 나머지 사운드 간의 유사도를 결정하고 반복 매칭 관계를 결정할 수 있다. 이를 위하여, 단계(330)은 일 세트의 서브 사운드의 나머지 서브 사운드 중 타겟 서브 사운드와 파형이 동일한 서브 사운드를 타겟 서브 사운드에 대한 반복 매칭 관계에 추가하는 단계를 포함할 수 있다.
- [0113] 추가적으로 사운드 분석 장치(100)는 일 세트의 서브 사운드 중 선택한 타겟 서브 사운드와 또다른 세트의 서브 사운드 간의 유사도를 결정하고 타겟 서브 사운드의 반복 매칭 관계를 결정할 수 있다. 이를 위하여, 단계(330)은 또다른 세트의 서브 사운드 중에서 타겟 서브 사운드와 파형이 동일한 서브 사운드를 타겟 서브 사운드에 대한 반복 매칭 관계에 추가하는 단계를 포함할 수 있다.
- [0114] 일 예에서, 또다른 세트의 서브 사운드는 단계(320)에서 일 세트의 서브 사운드로 분할된 입력 사운드와 동일한

입력 사운드를 분할하여 생성한 것이다. 일 예에서, 또다른 세트의 서브 사운드는 단계(320)에서 일 세트의 서브 사운드로 분할된 입력 사운드를 획득한 타겟 장소에서 이전에 획득된 또다른 입력 사운드를 분할하여 생성한 것이다. 이하에서 도 6을 참조하여 예를 들어 설명한다.

- [0115] 도 6은 일 실시예에 따른 반복 매칭 관계를 설명하기 위한 도면이다.
- [0116] 도 6(a)은 입력 사운드를 분할한 일 세트의 서브 사운드 내의 반복 매칭 관계를 보여준다. 타겟 서브 사운드가 서브 사운드(S_1)인 경우 서브 사운드(S_1)은 서브 사운드(S_7)와 매칭되지만, 서브 사운드(S_2)와는 매칭되지 않으므로, 사운드 분석 장치(100)는 S_1 - S_7 을 S_1 의 반복 매칭 관계로 결정한다. 타겟 서브 사운드가 서브 사운드(S_2)인 경우 서브 사운드(S_2)는 매칭되는 서브 사운드가 없으므로 반복 매칭 관계를 가지지 않게 된다.
- [0117] 도 6(b)은 동일한 입력 사운드로부터 분할 시작 지점을 달리하여 생성된 제 1 세트의 서브 사운드 및 제 2 세트의 서브 사운드에서 반복 매칭 관계를 보여준다. 사운드 분석 장치(100)는 제 1 세트의 서브 사운드의 서브 사운드(S_1)가 타겟 서브 사운드인 경우, 제 2 세트의 서브 사운드(S_2)와 매칭되므로 제 2 세트의 서브 사운드(S_2)를 타겟 서브 사운드(S_1)의 반복 매칭 관계에 추가한다. 사운드 분석 장치(100)는 제 1 세트의 서브 사운드의 서브 사운드(S_2)가 타겟 서브 사운드인 경우, 제 2 세트의 서브 사운드(S_3)와 매칭되므로 제 2 세트의 서브 사운드(S_3)를 타겟 서브 사운드(S_2)의 반복 매칭 관계에 추가한다.
- [0118] 도 6(c)에서 6(b)와 동일한 입력 사운드로부터 시간 간격을 달리하여 생성된 제 1 세트의 서브 사운드 및 제 2 세트의 서브 사운드에서 반복 매칭 관계를 보여준다. 사운드 분석 장치(100)는 제 1 세트의 서브 사운드의 서브 사운드(S_1)가 타겟 서브 사운드인 경우, 제 2 세트의 서브 사운드(S_2)와 매칭되므로 제 2 세트의 서브 사운드(S_2)를 타겟 서브 사운드(S_1)의 반복 매칭 관계에 추가한다. 도 6(c)는 도 6(b)보다 서브 사운드의 길이가 길기 때문에 더 긴 타임 슬롯에 대하여 반복 매칭 관계가 도출될 수 있다.
- [0119] 단계(340)에서 사운드 분석 장치(100)는 단계(330)에서 결정한 반복 매칭 관계에 기반하여 타겟 공간에서 반복적으로 발생하는 반복 사운드 패턴을 추출할 수 있다.
- [0120] 사운드 분석 장치(100)는 단계(340)에서 타겟 서브 사운드의 반복 매칭 관계의 반복 주기에 기반하여 타겟 서브 사운드가 반복 사운드 패턴으로 결정한다.
- [0121] 서브 사운드의 반복 주기는 반복 매칭 관계의 반복 위치에 기반하여 결정될 수 있다. 예를 들어 제 1 서브 사운드와 제 2 서브 사운드가 반복 매칭 관계에 있는 경우, 반복 주기는 제 1 서브 사운드의 타임 슬롯(예를 들어 2시-3시)과 제 2 서브 사운드의 타임 슬롯(예를 들어 5시-6시) 간의 거리(타임 슬롯 간의 차, 즉 3시간)로 결정될 수 있다.
- [0122] 타겟 서브 사운드가 속한 서브 사운드 세트 내에서 타겟 서브 사운드가 반복 매칭 관계를 가지는 경우, 사운드 분석 장치(100)는 단계(340)에서 다음의 단계들을 수행할 수 있다. 즉, 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 다음의 단계들을 수행한다. 즉, 사운드 분석 장치(100)는,
- [0123] (단계 1-1) 일 세트의 서브 사운드에서 반복 매칭 관계를 가지는 타겟 서브 사운드를 선택하는 단계,
- [0124] (단계 1-2) 일 세트의 서브 사운드에서 타겟 서브 사운드의 제 1 반복 주기를 결정하는 단계,
- [0125] (단계 1-3) 타겟 공간에서 이전에 획득된 또다른 세트의 서브 사운드의 반복 매칭 관계에 기반하여 또다른 세트의 서브 사운드에서 타겟 서브 사운드의 제 2 반복 주기를 결정하는 단계, 및
- [0126] (단계 1-4) 제 1 반복 주기와 제 2 반복 주기의 차이가 임계값 이하인 경우에 타겟 서브 사운드를 반복 사운드 패턴으로 추출하는 단계를 수행할 수 있다. 여기서 임계값은 타겟 서브 사운드의 시간 간격 이하로 설정될 수 있다. 예를 들어 타겟 서브 사운드가 1시간 구간이면 임계값은 1시간이하로 설정될 수 있다. 예를 들어 임계값은 0으로 설정될 수 있다. 일 예에서 임계값은 사운드 환경에서 발생하는 사운드의 종류, 크기 및 반복성 등에 따라 다르게 설정될 수 있고 조정가능하다. 타겟 서브 사운드가 속한 서브 사운드 세트 내에서 타겟 서브 사운드가 반복 매칭 관계를 가지는 않는 경우, 사운드 분석 장치(100)는 단계(340)에서 다음의 단계들을 수행할 수 있다. 즉, 사운드 분석 장치(100)는 프로세서(160)의 제어 하에 다음의 단계들을 수행한다. 즉, 사운드 분석 장치(100)는,
- [0127] (단계 2-1) 일 세트의 서브 사운드에서 반복 매칭 관계가 없는 타겟 서브 사운드를 선택하는 단계,

- [0128] (단계 2-2) 타겟 공간에서 이전에 획득된 또다른 세트의 서브 사운드에서 타겟 서브 사운드와 매칭되는 서브 사운드를 찾는 단계, 및
- [0129] (단계 2-3) 타겟 서브 사운드의 타임 슬롯과 매칭되는 서브 사운드의 타임 슬롯의 차이가 임계값 이하인 경우에 타겟 서브 사운드를 반복 사운드 패턴으로 추출하는 단계를 수행할 수 있다.
- [0130] 여기서 임계값은 타겟 서브 사운드의 시간 간격 이하로 설정될 수 있다. 예를 들어 타겟 서브 사운드가 1시간 구간이면 임계값은 1시간이하로 설정될 수 있다. 예를 들어 임계값은 0으로 설정될 수 있다. 일 예에서 임계값은 사운드 환경에서 발생하는 사운드의 종류, 크기 및 반복성 등에 따라 다르게 설정될 수 있고 조정가능하다.
- [0131] 이하에서, 도 7을 참조하여 반복 사운드 패턴 추출에 대하여 예시적으로 설명한다.
- [0132] 도 7은 일 실시예에 따른 반복 사운드 패턴을 설명하기 위한 도면이다.
- [0133] 도 7(a)는 타겟 장소에서 제 1 일에 24시간 동안 획득된 제 1 입력 사운드이고, 도 7(b)는 동일한 타겟 장소에서 제 1 일과 다른 제 2 일에 24시간 동안 획득된 제 2 입력 사운드를 나타낸다.
- [0134] 제 1 입력 사운드의 서브 사운드(S_1)은 동일자의 서브 사운드(S_7) 및 제 2 입력 사운드의 서브 사운드(S_1)와 서브 사운드(S_7)와 반복 매칭 관계에 있다고 가정한다. 또한, 제 1 입력 사운드의 서브 사운드(S_2)는 제 1 입력 사운드 내에서는 반복 매칭 관계가 없고, 제 2 입력 사운드의 서브 사운드(S_2)와 반복 매칭 관계에 있다고 가정한다.
- [0135] 도 7(a)의 제 1 입력 사운드의 서브 사운드(S_1)이 타겟 서브 사운드인 경우, 타겟 서브 사운드가 해당 세트의 서브 사운드 내에서 반복 매칭 관계를 가지는 경우이므로 전술한 (단계 1-1) 내지 (단계 1-4)를 수행하여 사운드 패턴(S_1)을 반복 사운드 패턴으로 추출할 것인지를 결정할 수 있다.
- [0136] 구체적으로, 사운드 분석 장치(100)는 단계(1-1)에서 타겟 서브 사운드(S_1)을 선택하고, 단계(1-2)에서 제 1 입력 사운드에서의 타겟 서브 사운드(S_1)의 제 1 반복 주기를 결정한다. 예를 들어, 서브 사운드(S_1)가 제 1 입력 사운드에서 0시에서 1시의 타임 슬롯이고, 서브 사운드(S_7)가 6시에서 7시의 타임 슬롯에 해당한다면, 제 1 반복 주기는 5시간이다. 사운드 분석 장치(100)는 마찬가지로 단계(1-3)에서 제 2 입력 사운드에서 타겟 서브 사운드(S_1)의 제 2 반복 주기를 5시간으로 결정할 수 있다. 사운드 분석 장치(100)는 단계(1-4)에서 제 1 반복 주기와 제 2 반복 주기의 차이가 0이므로 타겟 서브 사운드(S_1)을 제 1 입력 사운드의 반복 사운드 패턴으로 결정할 수 있다. 즉, 제 1 일에 반복된 주기와 제 2 일에 반복된 주기가 동일하거나 임계값 이하인 서브 사운드를 타겟 공간의 반복 사운드 패턴으로 결정할 수 있다.
- [0137] 사운드 분석 장치(100)는 서브 사운드의 타임 슬롯 대신에 서브 사운드에서 발생한 사운드 이벤트(e)의 발생 시점의 반복 주기(d)의 차이가 임계값(α) 이하인 지에 따라서 다음의 수학적 식 1을 이용하여 반복 사운드 패턴을 결정할 수 있다.
- [0138] (수학적 식 1)
- $$|d_2(e_1, e'_1) - d_1(e_1, e'_1)| < \alpha$$
- [0139]
- [0140] 예를 들어, 타겟 서브 사운드(S_1)의 사운드 이벤트(e_1)의 발생 시점과 타겟 서브 사운드와 반복 매칭 관계에 있는 서브 사운드(S_7)의 사운드 이벤트(e'_1)의 발생 시점 간의 차이를 반복 주기(d)로 하여 단계(1-2) 내지 단계(1-4)를 수행할 수 있다.
- [0141] 도 7(a)의 제 1 입력 사운드의 서브 사운드(S_2)가 타겟 서브 사운드인 경우, 타겟 서브 사운드가 해당 세트의 서브 사운드 내에서 반복 매칭 관계를 가지지 않는 경우이므로 전술한 (단계 2-1) 내지 (단계 2-3)를 수행하여 사운드 패턴(S_2)을 반복 사운드 패턴으로 추출할 것인지를 결정할 수 있다.
- [0142] 구체적으로, 사운드 분석 장치(100)는 단계(2-1)에서 타겟 서브 사운드(S_2)를 선택하고, 단계(2-2)에서 타겟 서브 사운드(S_2)와 매칭되는 제 2 입력 사운드의 서브 사운드(S_2)를 찾는다. 사운드 분석 장치(100)는 단계(2-3)에

서 타겟 서브 사운드(S_2)의 타임 슬롯과 매칭되는 제 2 입력 사운드의 서브 사운드(S_2)의 타임 슬롯의 차이를 계산한다. 예를 들어, 타겟 서브 사운드(S_2)가 제 1 입력 사운드에서 3시에서 4시의 타임 슬롯이고, 서브 사운드(S_2)가 제 2 입력 사운드에서 3시에서 4시의 타임 슬롯에 해당한다면, 타임 슬롯 차이는 0이다. 사운드 분석 장치(100)는 단계(2-3)에서 타임 슬롯의 차이가 0이므로 타겟 서브 사운드(S_2)를 제 1 입력 사운드의 반복 사운드 패턴으로 결정할 수 있다. 즉, 제 1 일과 제 2 일에 동일 시간대에 반복된 서브 사운드를 타겟 장소의 반복 사운드 패턴으로 추출할 수 있다.

[0143] 사운드 분석 장치(100)는 서브 사운드의 타임 슬롯 대신에 서브 사운드에서 발생한 사운드 이벤트(e)의 발생 시점(p)의 차이가 임계값(β) 이하인 지에 따라서 다음의 수학식에 따라 반복 사운드 패턴을 결정할 수 있다.

[0144] (수학식 2-1)

$$p_2(e_2) = |s - e_2|$$

[0145]

[0146] (수학식 2-2)

$$p_1(e_2) = |s - e_2|$$

[0147]

[0148] (수학식 2-3)

$$|p_2(e_2) - p_1(e_2)| < \beta$$

[0149]

[0150] 예를 들어, 타겟 서브 사운드(S_2)의 사운드 이벤트(e_1)의 발생 시점(p_1)과 타겟 서브 사운드와 반복 매칭 관계에 있는 제 2 입력 사운드의 서브 사운드(S_2)의 사운드 이벤트(e_2)의 발생 시점(p_2) 간의 차이가 임계값(β) 이하인 지에 따라 타겟 서브 사운드(S_2)를 타겟 장소의 반복 사운드 패턴으로 추출할 수 있다.

[0151] 단계(350)에서 사운드 분석 장치(100)는 단계(340)에서 생성한 반복 사운드 패턴에 기반하여 타겟 공간에 대한 사운드 학습 모델을 훈련할 수 있다.

[0152] 단계(350)에서 사운드 학습 장치(100)는, 단계(340)에서 타겟 공간에서 추출한 반복 사운드 패턴을 사운드 학습 모델에 학습 데이터로 제공할 수 있다.

[0153] 사운드 학습 장치(100)는 반복 사운드 패턴으로부터 특징을 추출하기 위하여 딥 오토 인코더(deep auto-encoder)에 기반한 사운드 학습 모델을 학습시킬 수 있다.

[0154] 오토 인코더는 입력 자체를 출력으로 재현하는 것을 목표로 하는 신경망이다. 오토 인코더는 입력층, 적어도 하나의 은닉층 및 출력층을 포함한다.

[0155] 이 경우 은닉 계층의 노드 수가 입력 계층의 노드 수보다 적으므로 데이터의 차원이 줄어들게 되며, 이에 따라 압축 또는 인코딩이 수행되게 된다.

[0156] 또한 은닉 계층에서 출력한 데이터는 출력 계층으로 들어간다. 이 경우 출력 계층의 노드 수는 은닉 계층의 노드 수보다 많으므로, 데이터의 차원이 늘어나게 되며, 이에 따라 압축 해제 또는 디코딩이 수행되게 된다.

[0157] 한편 오토 인코더는 학습을 통해 뉴런의 연결 강도를 조절함으로써 입력 데이터가 은닉층 데이터로 표현된다. 은닉층에서는 입력층보다 적은 수의 뉴런으로 정보를 표현하는데 입력 데이터를 출력으로 재현할 수 있다는 것은, 은닉층이 입력 데이터로부터 숨은 패턴을 발견하여 표현했다는 것을 의미할 수 있다.

[0158] 일 예에서, 사운드 학습 장치(100)는 타겟 장소별로 동일한 길이의 시간 간격에 따라 분할된 서브 사운드 세트를 사운드 학습 모델에 입력 데이터로 제공할 수 있다. 일 예에서, 사운드 학습 장치(100)는 타겟 장소별로 사운드 학습 모델을 생성할 수 있다.

[0159] 실시예에 따른 사운드 분석 방법은 추가적으로 단계(350)에서 훈련된 사운드 학습 모델에 기반하여 타겟 공간에서 획득한 타겟 사운드의 반복 사운드 패턴을 결정하는 단계를 더 포함할 수 있다. 이에 대하여는 도 8을 참조

하여 후술한다.

- [0160] 도 8은 일 실시예에 따른 타겟 사운드 분석 과정의 흐름도이다.
- [0161] 실시예에 따른 사운드 분석 방법은 도 3을 참조하여 훈련된 사운드 학습 모델을 이용하여 타겟 공간에서 획득한 타겟 사운드로부터 타겟 사운드의 반복 사운드 패턴을 결정하는 단계를 더 포함할 수 있다.
- [0162] 사운드 분석 장치(100)는 도 3을 참조하여 전술한 사운드 분석 방법에 의해 훈련된 사운드 학습 모델을 이용하여 타겟 공간에서 획득한 타겟 사운드의 반복 사운드 패턴을 결정할 수 있다.
- [0163] 타겟 사운드의 반복 사운드 패턴을 결정하는 단계는, 타겟 공간에서 타겟 사운드를 획득하는 단계(810), 타겟 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 타겟 서브 사운드를 생성하는 단계(820) 및 사운드 학습 모델에 기반하여 일 세트의 타겟 서브 사운드로부터 타겟 사운드의 반복 사운드 패턴을 결정하는 단계(830)를 포함할 수 있다.
- [0164] 단계(810)에서 사운드 분석 장치(100)는 타겟 공간에서 타겟 사운드를 획득할 수 있다. 단계(810)에서의 타겟 사운드 획득은 도 3을 참조하여 전술한 단계(310)의 입력 사운드 획득 과정을 그대로 적용할 수 있다.
- [0165] 단계(820)에서 사운드 분석 장치(100)는 단계(810)에서 획득한 타겟 사운드를 소정의 시간 간격에 따라 분할하여 일 세트의 타겟 서브 사운드를 생성할 수 있다. 단계(820)은 도 3을 참조하여 전술한 단계(320)의 과정을 그대로 적용할 수 있다.
- [0166] 단계(830)에서 사운드 분석 장치(100)는 사운드 학습 모델에 기반하여 단계(820)에서 생성한 일 세트의 타겟 서브 사운드로부터 타겟 사운드의 반복 사운드 패턴을 결정할 수 있다.
- [0167] 단계(830)에서 사운드 분석 장치(100)는 타겟 사운드 및 타겟 서브 사운드의 길이에 따라 적합한 특징 추출 모델을 선택할 수 있다. 단계(830)에서 사운드 분석 장치(100)는 오토 인코더에 기반한 사운드 학습 모델의 가중치를 통해 각 타겟 서브 사운드의 특징을 추출하고, 이로써 각 타겟 서브 사운드가 반복 사운드 패턴인지 여부를 판단할 수 있다.
- [0168] 추가적으로 타겟 사운드의 반복 사운드 패턴을 결정하는 단계는, 반복 사운드 패턴을 타겟 서브 사운드로부터 제거하는 단계(840)를 더 포함할 수 있다.
- [0169] 단계(840)에서 사운드 분석 장치(100)는 타겟 공간의 반복 사운드 패턴인 것으로 결정된 타겟 서브 사운드를 타겟 공간의 노이즈로 결정할 수 있다. 사운드 분석 장치(100)는 타겟 서브 사운드에서 노이즈를 제거하는 전처리 과정에서 타겟 공간의 반복 사운드 패턴인 것으로 결정된 타겟 서브 사운드를 타겟 사운드로부터 제거할 수 있다. 이로써, 타겟 공간의 사운드 환경에 특화된 노이즈의 제공이 가능하여 음성 인식 성능이 제고된다.
- [0170] 이상 설명된 본 발명에 따른 실시예는 컴퓨터 상에서 다양한 구성요소를 통하여 실행될 수 있는 컴퓨터 프로그램의 형태로 구현될 수 있으며, 이와 같은 컴퓨터 프로그램은 컴퓨터로 판독 가능한 매체에 기록될 수 있다. 이때, 매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등과 같은, 프로그램 명령어를 저장하고 실행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다.
- [0171] 한편, 상기 컴퓨터 프로그램은 본 발명을 위하여 특별히 설계되고 구성된 것이거나 컴퓨터 소프트웨어 분야의 당업자에게 공지되어 사용 가능한 것일 수 있다. 컴퓨터 프로그램의 예에는, 컴파일러에 의하여 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용하여 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드도 포함될 수 있다.
- [0172] 본 발명의 명세서(특히 청구범위에서)에서 "상기"의 용어 및 이와 유사한 지시 용어의 사용은 단수 및 복수 모두에 해당하는 것일 수 있다. 또한, 본 발명에서 범위(range)를 기재한 경우 상기 범위에 속하는 개별적인 값을 적용한 발명을 포함하는 것으로서(이에 반하는 기재가 없다면), 발명의 상세한 설명에 상기 범위를 구성하는 각 개별적인 값을 기재한 것과 같다.
- [0173] 본 발명에 따른 방법을 구성하는 단계들에 대하여 명백하게 순서를 기재하거나 반하는 기재가 없다면, 상기 단계들은 적당한 순서로 행해질 수 있다. 반드시 상기 단계들의 기재 순서에 따라 본 발명이 한정되는 것은 아니다. 본 발명에서 모든 예들 또는 예시적인 용어(예들 들어, 등등)의 사용은 단순히 본 발명을 상세히 설명하기 위한 것으로서 특허청구범위에 의해 한정되지 않는 이상 상기 예들 또는 예시적인 용어로 인해 본 발명의 범위가 한정되는 것은 아니다. 또한, 당업자는 다양한 수정, 조합 및 변경이 부가된 청구의 범위 또는 그 균등물의

범주 내에서 설계 조건 및 팩터에 따라 구성될 수 있음을 알 수 있다.

[0174] 따라서, 본 발명의 사상은 상기 설명된 실시 예에 국한되어 정해져서는 아니 되며, 후술하는 청구의 범위뿐만 아니라 이 청구의 범위와 균등한 또는 이로부터 등가적으로 변경된 모든 범위는 본 발명의 사상의 범주에 속한다고 할 것이다.

[0175] 앞에서, 본 발명의 특정한 실시예가 설명되고 도시되었지만 본 발명은 기재된 실시예에 한정되는 것이 아니고, 이 기술 분야에서 통상의 지식을 가진 자는 본 발명의 사상 및 범위를 벗어나지 않고서 다른 구체적인 실시예로 다양하게 수정 및 변형할 수 있음을 이해할 수 있을 것이다. 따라서, 본 발명의 범위는 설명된 실시예에 의하여 정하여 질 것이 아니고 청구범위에 기재된 기술적 사상에 의해 정하여져야 할 것이다.

부호의 설명

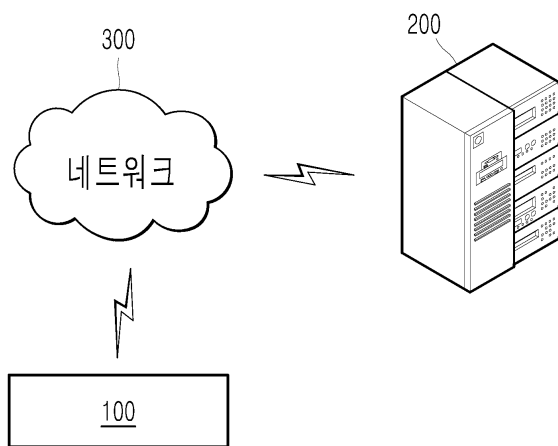
[0176] 100: 사운드 분석 장치

200: 서버

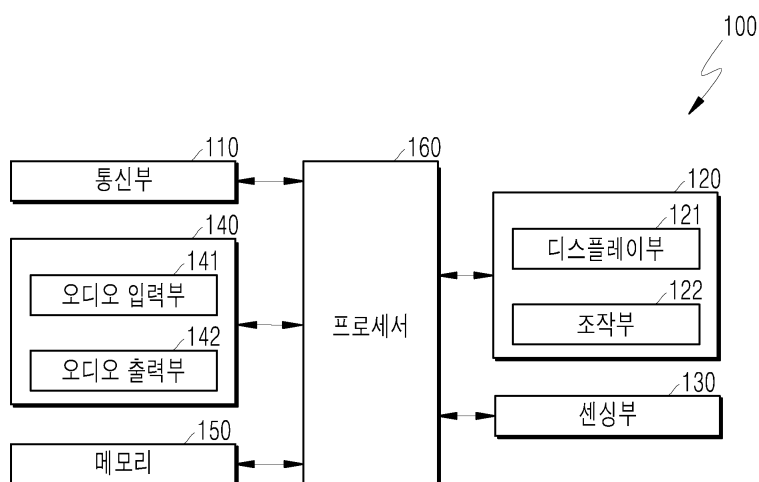
300: 네트워크

도면

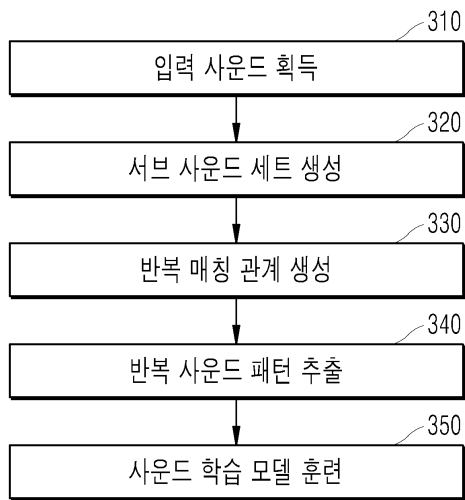
도면1



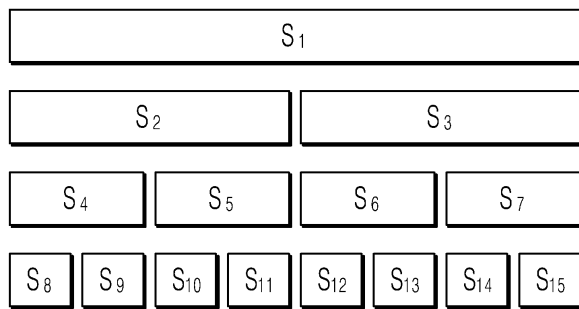
도면2



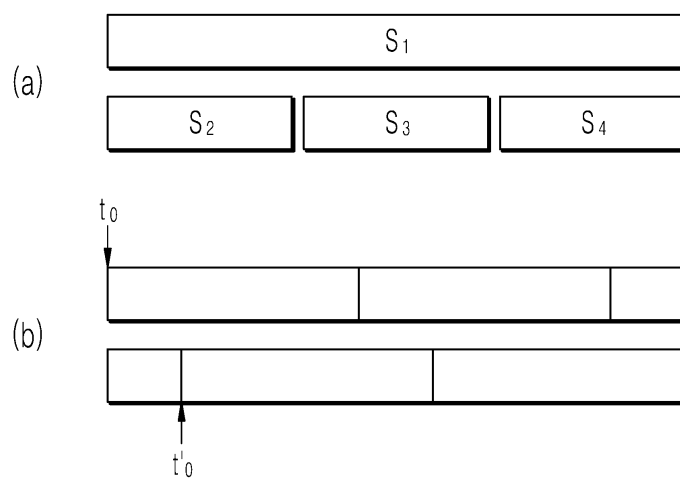
도면3



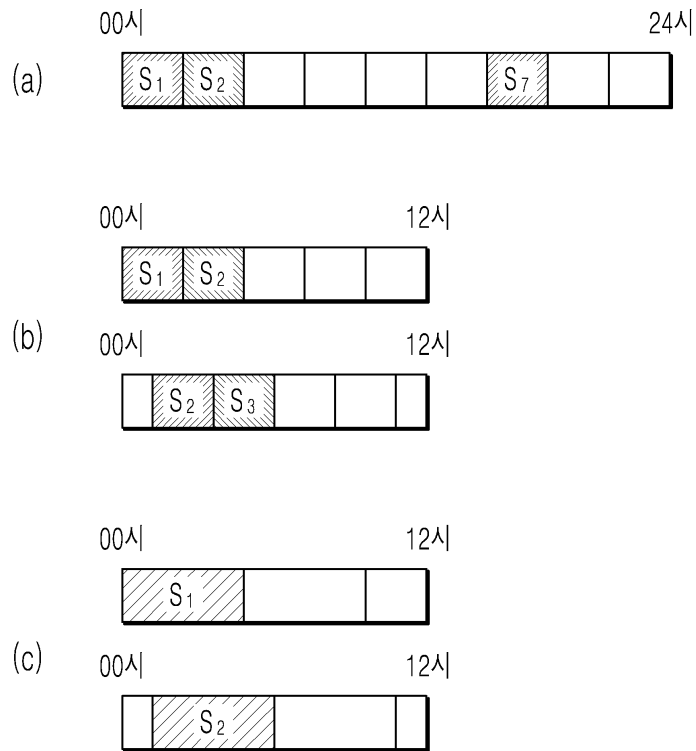
도면4



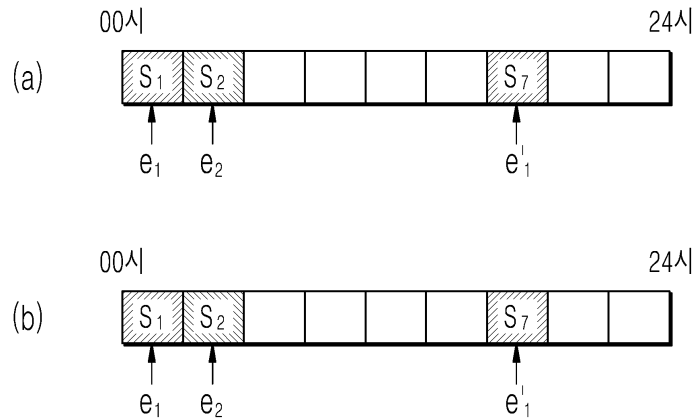
도면5



도면6



도면7



도면8

