(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0110513
(43) 공개일자 2023년07월24일

(51) 국제특허분류(Int. Cl.)
*G10L 25/51* (2013.01) *G06F 16/65* (2019.01)
*G06N 20/00* (2019.01) *G06N 5/04* (2023.01)
*G10L 25/30* (2013.01) *H04R 1/08* (2006.01)
(52) CPC특허분류
*G10L 25/51* (2013.01)
*G06F 16/65* (2019.01)
(21) 출원번호 10-2023-7016431
(22) 출원일자(국제) 2021년11월19일
심사청구일자 없음
(85) 번역문제출일자 2023년05월15일
(86) 국제출원번호 PCT/US2021/072520
(87) 국제공개번호 WO 2022/115838
국제공개일자 2022년06월02일
(30) 우선권주장
17/102,724 2020년11월24일 미국(US)

(71) 출원인
**퀄컴 인코포레이티드**
미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
(72) 발명자
**사키, 파테메**
미국 92121 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
**궈, 인이**
미국 92121 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
**비저, 에릭**
미국 92121 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
(74) 대리인
**특허법인 남앤남**

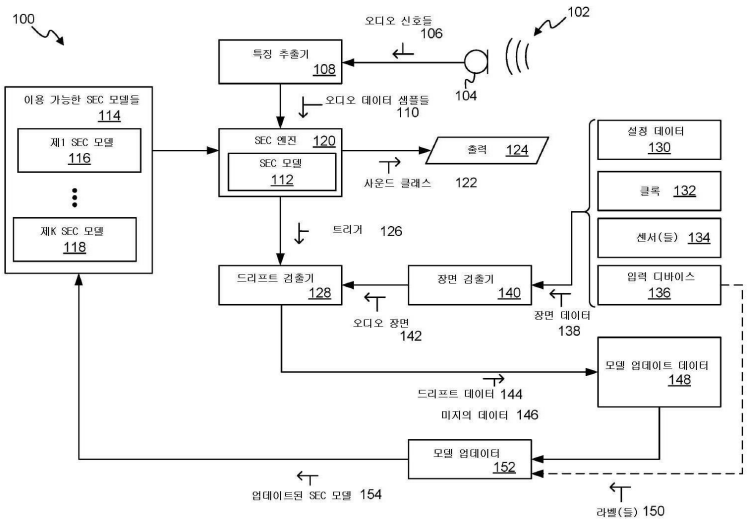
전체 청구항 수 : 총 30 항

(54) 발명의 명칭 **적응형 사운드 이벤트 분류**

## (57) 요 약

디바이스는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 또한, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 구성된다. 하나 이상의 프로세서들은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하도록 추가로 구성된다. 하나 이상의 프로세서들은 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하도록 구성된다.

## 대 표 도



(52) CPC특허분류
*G06N 20/00* (2021.08)
*G06N 5/04* (2023.01)
*G10L 25/30* (2013.01)
*H04R 1/08* (2013.01)

# 명 세 서

## 청구범위

### 청구항 1

하나 이상의 프로세서들을 포함하며,

상기 하나 이상의 프로세서들은:

오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하도록;

상기 오디오 데이터 샘플들에 대한 응답으로 상기 사운드 이벤트 분류 모델의 출력에 기반하여, 상기 오디오 데이터 샘플들의 사운드 클래스가 상기 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록;

상기 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하도록; 그리고

상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 상기 오디오 장면에 대응한다는 결정에 기반하여, 상기 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하도록 구성되는, 디바이스.

### 청구항 2

제1 항에 있어서,

상기 하나 이상의 프로세서들에 커플링되고, 상기 오디오 데이터 샘플들에 대응하는 오디오 데이터를 캡처하도록 구성된 마이크로폰을 더 포함하는, 디바이스.

### 청구항 3

제1 항에 있어서,

상기 하나 이상의 프로세서들에 커플링되고 복수의 사운드 이벤트 분류 모델들을 저장하도록 구성된 메모리를 더 포함하고, 상기 하나 이상의 프로세서들은 상기 복수의 사운드 이벤트 분류 모델들 중에서 상기 사운드 이벤트 분류 모델을 선택하도록 구성되는, 디바이스.

### 청구항 4

제3 항에 있어서,

상기 오디오 데이터 샘플들과 연관된 센서 데이터를 생성하도록 구성된 하나 이상의 센서들을 더 포함하고, 상기 하나 이상의 프로세서들은 상기 센서 데이터에 기반하여 상기 사운드 이벤트 분류 모델을 선택하도록 구성되는, 디바이스.

### 청구항 5

제4 항에 있어서,

상기 하나 이상의 센서들은 카메라 및 포지션 센서를 포함하는, 디바이스.

### 청구항 6

제3 항에 있어서,

상기 오디오 장면을 식별하는 입력을 수신하도록 구성되는 하나 이상의 입력 디바이스들을 더 포함하고, 상기 하나 이상의 프로세서들은 상기 오디오 장면에 기반하여 상기 사운드 이벤트 분류 모델을 선택하도록 구성되는, 디바이스.

### 청구항 7

제3 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 오디오 데이터 샘플들이 수신되는 경우에 기반하여 상기 사운드 이벤트 분류 모델을 선택하도록 구성되는, 디바이스.

**청구항 8**

제3 항에 있어서,

상기 메모리는 하나 이상의 디바이스 설정들을 나타내는 설정 데이터를 추가로 저장하고, 그리고 상기 하나 이상의 프로세서들은 상기 설정 데이터에 기반하여 상기 사운드 이벤트 분류 모델을 선택하도록 구성되는, 디바이스.

**청구항 9**

제1 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 사운드 클래스가 인식되었다는 결정에 기반하여, 상기 오디오 데이터 샘플들과 연관된 상기 사운드 클래스를 표시하는 출력을 생성하도록 추가로 구성되는, 디바이스.

**청구항 10**

제1 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 상기 오디오 장면에 대응하지 않는다는 결정에 기반하여, 상기 오디오 데이터 샘플들에 대응하는 오디오 데이터를 새로운 사운드 이벤트 분류 모델에 대한 트레이닝 데이터로서 저장하도록 추가로 구성되는, 디바이스.

**청구항 11**

제1 항에 있어서,

상기 사운드 이벤트 분류 모델은 상기 출력과 연관된 신뢰도 메트릭을 생성하도록 추가로 구성되고, 그리고 상기 하나 이상의 프로세서들은, 상기 사운드 클래스가 상기 신뢰도 메트릭에 기반하여 상기 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 구성되는, 디바이스.

**청구항 12**

제1 항에 있어서,

상기 하나 이상의 프로세서들은 추가로, 상기 모델 업데이트 데이터에 기반하여 상기 사운드 이벤트 분류 모델을 업데이트하도록 구성되는, 디바이스.

**청구항 13**

제1 항에 있어서,

상기 오디오 장면을 식별하는 입력을 수신하도록 구성되는 하나 이상의 입력 디바이스들을 더 포함하고, 상기 하나 이상의 프로세서들은 상기 사운드 이벤트 분류 모델이 상기 입력에 기반하여 상기 오디오 장면에 대응하는지 여부를 결정하도록 구성되는, 디바이스.

**청구항 14**

제1 항에 있어서,

상기 오디오 데이터 샘플들과 연관된 센서 데이터를 생성하도록 구성된 하나 이상의 센서들을 더 포함하고, 상기 하나 이상의 프로세서들은 상기 사운드 이벤트 분류 모델이 상기 센서 데이터에 기반하여 상기 오디오 장면에 대응하는지 여부를 결정하도록 구성되는, 디바이스.

**청구항 15**

제14 항에 있어서,

상기 하나 이상의 센서들은 카메라 및 포지션 센서를 포함하는, 디바이스.

**청구항 16**

제14 항에 있어서,

상기 하나 이상의 프로세서들은 추가로, 상기 오디오 데이터 샘플들과 연관된 타임스탬프에 기반하여, 상기 사운드 이벤트 분류 모델이 상기 오디오 장면에 대응하는지 여부를 결정하도록 구성되는, 디바이스.

**청구항 17**

제1 항에 있어서,

상기 사운드 이벤트 분류 모델은 특정 사운드 클래스를 인식하도록 트레이닝되고, 상기 모델 업데이트 데이터는, 상기 사운드 이벤트 분류 모델이 특정 사운드 클래스에 대응하는 것으로 인식하도록 트레이닝되지 않은, 상기 특정 사운드 클래스 내의 사운드의 특성들의 변동을 표현하는 드리프트 데이터를 포함하는, 디바이스.

**청구항 18**

제1 항에 있어서,

상기 하나 이상의 프로세서들은 모바일 컴퓨팅 디바이스, 차량, 웨어러블 디바이스, 증강 현실 헤드셋, 혼합 현실 헤드셋 또는 가상 현실 헤드셋 내에 통합되는, 디바이스.

**청구항 19**

제1 항에 있어서,

상기 하나 이상의 프로세서들은 집적 회로에 포함되는, 디바이스.

**청구항 20**

하나 이상의 프로세서들이, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 입력으로서 제공하는 단계;

상기 오디오 데이터 샘플들에 대한 응답으로 상기 사운드 이벤트 분류 모델의 출력에 기반하여, 상기 오디오 데이터 샘플들의 사운드 클래스가 상기 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 상기 하나 이상의 프로세서들이 결정하는 단계;

상기 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 상기 하나 이상의 프로세서들이 결정하는 단계; 및

상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 상기 오디오 장면에 대응한다는 결정에 기반하여, 상기 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 상기 하나 이상의 프로세서들이 저장하는 단계를 포함하는, 방법.

**청구항 21**

제20 항에 있어서,

상기 하나 이상의 프로세서들에 커플링된 메모리에 저장된 복수의 사운드 이벤트 분류 모델들 중에서 상기 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함하는, 방법.

**청구항 22**

제21 항에 있어서,

상기 사운드 이벤트 분류 모델은 사용자 입력, 설정 데이터, 로케이션 데이터, 이미지 데이터, 비디오 데이터, 상기 오디오 데이터 샘플들과 연관된 타임스탬프 또는 이들의 조합에 기반하여 선택되는, 방법.

**청구항 23**

제20 항에 있어서,

상기 사운드 이벤트 분류 모델이 상기 오디오 장면에 대응하는지 여부의 결정은, 상기 사운드 이벤트 분류 모델에 의해 생성된 신뢰도 메트릭, 사용자 입력, 설정 데이터, 로케이션 데이터, 이미지 데이터, 비디오 데이터,

오디오 데이터 샘플들과 연관된 타임스탬프 또는 이들의 조합에 기반하는, 방법.

**청구항 24**

제20 항에 있어서,

상기 모델 업데이트 데이터를 저장한 후에,

모델 업데이트 데이터의 임계량이 누적되었는지 여부를 결정하는 단계, 및

모델 업데이트 데이터의 임계량이 누적되었다는 결정에 기반하여, 누적된 모델 업데이트 데이터를 사용하여 상기 사운드 이벤트 분류 모델의 자동 업데이트를 개시하는 단계를 더 포함하는, 방법.

**청구항 25**

제24 항에 있어서,

상기 자동 업데이트 전에, 상기 사운드 이벤트 분류 모델은 특정 사운드 클래스의 다수의 변화들을 인식하도록 트레이닝되었고, 그리고 상기 자동 업데이트는 상기 사운드 이벤트 분류 모델이 상기 특정 사운드 클래스에 대응하는 것으로 상기 특정 사운드 클래스의 추가적인 변형을 인식할 수 있게 하도록 상기 사운드 이벤트 분류 모델을 수정하는, 방법.

**청구항 26**

오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단;

상기 사운드 이벤트 분류 모델의 출력에 기반하여, 상기 오디오 데이터 샘플들의 사운드 클래스가 상기 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단;

상기 사운드 클래스가 인식되지 않았다는 결정에 응답하여, 상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단; 및

상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 상기 오디오 장면에 대응한다는 결정에 응답하여, 상기 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하기 위한 수단을 포함하는, 디바이스.

**청구항 27**

제26 항에 있어서,

선택 기준에 기반하여 복수의 사운드 이벤트 분류 모델들 중에서 상기 사운드 이벤트 분류 모델을 선택하기 위한 수단을 더 포함하는, 디바이스.

**청구항 28**

제26 항에 있어서,

상기 모델 업데이트 데이터에 기반하여 상기 사운드 이벤트 분류 모델을 업데이트하기 위한 수단을 더 포함하는, 디바이스.

**청구항 29**

명령들을 저장하는 비일시적 컴퓨터 판독 가능 저장 매체로서,

상기 명령들은 프로세서로 하여금,

오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하게 하고;

상기 오디오 데이터 샘플들에 대한 응답으로 상기 사운드 이벤트 분류 모델의 출력에 기반하여, 상기 오디오 데이터 샘플들의 사운드 클래스가 상기 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하게 하고;

상기 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하게 하고; 그리고

상기 사운드 이벤트 분류 모델이 상기 오디오 데이터 샘플들과 연관된 상기 오디오 장면에 대응한다는 결정에 기반하여, 상기 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하게 하도록,

상기 프로세서에 의해 실행 가능한, 비일시적 컴퓨터 판독 가능 저장 매체.

**청구항 30**

제29 항에 있어서,

상기 명령들은 추가로, 상기 프로세서로 하여금, 상기 모델 업데이트 데이터에 기반하여 상기 사운드 이벤트 분류 모델을 업데이트하게 하는, 비일시적인 컴퓨터 판독 가능 저장 매체.

## *발명의 설명*

### *기 술 분 야*

[0001] [0001] 본 출원은 2020년 11월 24일에 출원된, 공동 소유의 미국 정규 특허 출원 번호 제17/102,724호를 우선권으로 주장하며, 상기 출원의 내용들은 그 전체가 인용에 의해 본원에 명시적으로 포함된다.

[0002] [0002] 본 개시내용은 일반적으로 적응형 사운드 이벤트 분류에 관한 것이다.

### *배 경 기 술*

[0003] [0003] 기술에서의 진보들은 더 작고 더 강력한 컴퓨팅 디바이스들을 도출해왔다. 예를 들어, 작고, 경량이고, 사용자들에 의해 소지하기 쉬운 모바일 및 스마트 폰들, 태블릿들 및 랩탑 컴퓨터들과 같은 무선 전화들을 포함하는 다양한 휴대용 개인 컴퓨팅 디바이스들이 현재 존재한다. 이러한 디바이스들은 무선 네트워크들을 통해 음성 및 데이터 패킷들을 통신할 수 있다. 추가로, 많은 이러한 디바이스들은 디지털 스틸 카메라, 디지털 비디오 카메라, 디지털 레코더 및 오디오 파일 플레이어와 같은 추가적인 기능을 통합한다. 또한, 이러한 디바이스들은 인터넷에 액세스하기 위해 사용될 수 있는 소프트웨어 애플리케이션들, 예를 들어, 웹 브라우저 애플리케이션을 포함하는 실행 가능한 명령들을 프로세싱할 수 있다. 이에 따라, 이러한 디바이스들은 예컨대, 오디오 신호에서 사운드 이벤트들(예컨대, 문을 두들기는 것, 자동차 경적 등)을 인식하려고 시도하는 SEC(Sound Event Classification) 시스템을 포함하는 상당한 컴퓨팅 능력들을 포함할 수 있다.

[0004] [0004] SEC 시스템은 일반적으로, 라벨링된 트레이닝 데이터에서 식별되는 특정 세트의 사운드들을 인식하기 위해 지도 머신 러닝 기법을 사용하여 트레이닝된다. 결과적으로, 각각의 SEC 시스템은 도메인 특정적인 경향이 있다(예컨대, 사운드들의 미리 결정된 세트를 분류할 수 있음). SEC 시스템이 트레이닝된 후에, 라벨링된 트레이닝 데이터에서 식별되지 않은 새로운 사운드 클래스들을 인식하도록 SEC 시스템을 업데이트하는 것은 어렵다. 부가적으로, SEC 시스템이 검출하도록 트레이닝된 일부 사운드 클래스들은 라벨링된 트레이닝 데이터에 표현된 것보다 더 많은 변형들을 갖는 사운드 이벤트들을 표현할 수 있다. 예시하자면, 라벨링된 트레이닝 데이터는 많은 상이한 도어벨들에 대한 오디오 데이터 샘플들을 포함할 수 있지만, 도어벨 사운드들의 기존의 변형들 모두를 포함할 가능성은 없다. SEC 시스템을 트레이닝시키는 데 사용된 트레이닝 데이터에서 표현되지 않은 새로운 사운드를 인식하기 위해 SEC 시스템을 재트레이닝하는 것은, 오리지널 트레이닝 데이터에 부가하여 새로운 사운드에 대한 예들을 포함하는 라벨링된 트레이닝 데이터의 새로운 세트를 사용하여 SEC 시스템을 완전히 재트레이닝하는 것을 수반할 수 있다. 결과적으로, (새로운 사운드 클래스에 대해서든 또는 기존의 사운드 클래스의 변형이든) 새로운 사운드를 인식하도록 SEC 시스템을 트레이닝시키는 것은 최신 SEC 시스템를 생성하는 것과 대략 동일한 컴퓨팅 자원들(예컨대, 프로세서 사이클들, 메모리 등)을 요구한다. 또한, 시간이 지나면서, 훨씬 더 많은 사운드들이 인식되도록 추가됨에 따라, SEC 시스템을 트레이닝시키기 위해 유지 및 사용되어야 하는 오디오 데이터 샘플들의 수가 다루기 어려워질 수 있다.

### *발명의 내용*

[0005] [0005] 특정 양상에서, 디바이스는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하도록 그리고 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력을 수신하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 또한, 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 구성된다. 하나 이상의 프로세서들은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디

오 장면에 대응하는지 여부를 결정하도록 추가로 구성된다. 하나 이상의 프로세서들은 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하도록 구성된다. 특정 양상에서, 방법은 하나 이상의 프로세서들이, 오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하는 단계를 포함한다. 방법은 또한, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 하나 이상의 프로세서들이 결정하는 단계를 포함한다. 방법은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 하나 이상의 프로세서들이 결정하는 단계를 더 포함한다. 방법은 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 하나 이상의 프로세서들이 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하는 단계를 포함한다.

[0006] [0006] 특정 양상에서, 디바이스는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단을 포함한다. 디바이스는 또한, 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단을 포함한다. 디바이스는, 사운드 클래스가 인식되지 않았다는 결정에 응답하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단을 더 포함한다. 디바이스는 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 응답하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하기 위한 수단을 포함한다.

[0007] [0007] 특정 양상에서, 비일시적 컴퓨터 판독 가능 저장 매체는 프로세서에 의해 실행될 때, 프로세서로 하여금 오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하게 하는 명령들을 포함한다. 프로세서에 의해 실행될 때, 명령들은 또한, 프로세서로 하여금, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하게 한다. 프로세서에 의해 실행될 때, 명령들은 추가로, 프로세서로 하여금 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하게 한다. 프로세서에 의해 실행될 때, 명령들은 또한, 프로세서로 하여금 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하게 한다.

[0008] [0008] 본 개시의 다른 양상들, 이점들 및 특징들은, 하기 섹션들, 즉, 도면의 간단한 설명, 상세한 설명 및 청구항들을 포함하는 전체 출원의 검토 후 명백해질 것이다.

## 도면의 간단한 설명

[0009] [0009] 도 1은 오디오 데이터 샘플들에 대한 응답으로 사운드 식별 데이터를 생성하도록 구성되고 사운드 이벤트 분류 모델을 업데이트하도록 구성된 디바이스의 예의 블록도이다.

[0010] [0010] 도 2는 특정 예에 따라 드리프트를 처리하기 위해 사운드 이벤트 분류 모델을 업데이트하는 양상들을 예시하는 도면이다.

[0011] [0011] 도 3은 특정 예에 따라 새로운 사운드 클래스들을 처리하기 위해 사운드 이벤트 분류 모델을 업데이트하는 양상들을 예시하는 도면이다.

[0012] [0012] 도 4는 도 1의 디바이스의 동작의 특정 예를 예시하는 도면이다.

[0013] [0013] 도 5는 도 1의 디바이스의 동작의 다른 특정 예를 예시하는 도면이다.

[0014] [0014] 도 6은 도 1의 디바이스의 특정 예를 예시하는 블록도이다.

[0015] [0015] 도 7은 도 1의 디바이스의 양상들을 통합하는 차량의 예시적인 예이다.

[0016] [0016] 도 8은 도 1의 디바이스의 양상들을 통합하는 가상 현실, 혼합 현실 또는 증강 현실 헤드셋을 예시한다.

[0017] [0017] 도 9는 도 1의 디바이스의 양상들을 통합하는 웨어러블 전자 디바이스를 예시한다.

[0018] [0018] 도 10은 도 1의 디바이스의 양상들을 통합하는 음성 제어 스피커 시스템을 예시한다.

[0019] [0019] 도 11은 도 1의 디바이스의 양상들을 통합하는 카메라를 예시한다.

[0020] 도 12는 도 1의 디바이스의 양상들을 통합하는 모바일 디바이스를 예시한다.

[0021] 도 13은 도 1의 디바이스의 양상들을 통합하는 항공(aerial) 디바이스를 예시한다.

[0022] 도 14는 도 1의 디바이스의 양상들을 통합하는 헤드셋을 예시한다.

[0023] 도 15는 도 1의 디바이스의 양상들을 통합하는 기기를 예시한다.

[0024] 도 16은 도 1의 디바이스의 동작 방법의 예를 예시하는 흐름도이다.

## *발명을 실시하기 위한 구체적인 내용*

[0010] [0025] 사운드 이벤트 분류 모델들은 머신-러닝 기법들을 사용하여 트레이닝될 수 있다. 예컨대, 뉴럴 네트워크는 역전파 또는 다른 머신-러닝 트레이닝 기법들을 사용하여 사운드 이벤트 분류기로서 트레이닝될 수 있다. 이러한 방식으로 트레이닝된 뉴럴 네트워크는 본원에서 "사운드 이벤트 분류 모델"로 지칭된다. 이러한 방식으로 트레이닝된 사운드 이벤트 분류 모델은, 휴대용 컴퓨팅 디바이스가 사운드 이벤트 분류 모델을 저장 및 사용하기에 (점유된 저장 공간 측면에서) 충분히 작고 (동작 동안 사용되는 컴퓨팅 자원 측면에서) 충분히 단순할 수 있다. 사운드 이벤트 분류 모델을 트레이닝시키는 프로세스는 사운드 이벤트 분류 모델을 사용하여 사운드 이벤트 분류를 수행하는 데 사용되는 것보다 상당히 더 많은 프로세싱 자원들을 사용한다. 부가적으로, 트레이닝 프로세스는, 사운드 이벤트 분류 모델이 트레이닝되어 검출할 각각의 사운드 클래스에 대한 많은 오디오 데이터 샘플들을 포함하는 큰 세트의 라벨링된 트레이닝 데이터를 사용한다. 휴대용 컴퓨팅 디바이스 또는 다른 자원 제한된 컴퓨팅 디바이스 상에서 사운드 이벤트 분류 모델을 처음부터 트레이닝하는 것은 메모리 활용 또는 다른 컴퓨팅 자원들의 관점에서 엄두를 내지 못할 수 있다. 결과적으로, 휴대용 컴퓨팅 디바이스 상에서 사운드 이벤트 분류 모델을 사용하기를 원하는 사용자는 자원 제약이 덜한 컴퓨팅 디바이스로부터 또는 사전-트레이닝된 사운드 이벤트 분류 모델들의 라이브러리로부터 휴대용 컴퓨팅 디바이스로 사전-트레이닝된 사운드 이벤트 분류 모델들을 다운로드하는 것으로 제한될 수 있다. 따라서, 사용자는 제한된 맞춤화 옵션들을 갖는다.

[0011] [0026] 개시된 시스템들 및 방법들은 사운드 이벤트 분류 모델들을 처음부터 트레이닝시키는 것보다 상당히 더 적은 컴퓨팅 자원들을 사용하는 방식으로 사운드 이벤트 분류 모델들을 업데이트하기 위해 전이 러닝 기법들을 사용한다. 특정 양상에 따르면, 전이 러닝 기법들은 사운드 클래스 내의 드리프트를 처리하기 위해 또는 새로운 사운드 클래스를 인식하기 위해 사운드 이벤트 분류 모델을 업데이트하는 데 사용될 수 있다. 이러한 맥락에서, "드리프트"는 사운드 클래스 내의 변동을 지칭한다. 예컨대, 사운드 이벤트 분류 모델은 사운드 클래스의 일부 예들을 인식할 수 있지만, 사운드 클래스의 다른 예들을 인식하지 못할 수 있다. 예시하자면, 자동차 경적음 사운드 클래스를 인식하도록 트레이닝된 사운드 이벤트 분류 모델은 많은 상이한 유형들의 자동차 경적들을 인식할 수 있지만, 자동차 경적들의 일부 예들을 인식하지 못할 수 있다. 음향 환경의 변동들로 인해 드리프트가 또한 발생할 수 있다. 예시하자면, 사운드 이벤트 분류 모델은 콘서트 홀에서 연주되는 베이스 드럼의 사운드를 인식하도록 트레이닝될 수 있지만, 베이스 드럼이 야외 퍼레이드에서 악단에 의해 연주되는 경우에는 베이스 드럼을 인식하지 못할 수 있다. 본 명세서에 개시된 전이 러닝 기법들은 이러한 드리프트를 처리하기 위해 사운드 이벤트 분류 모델을 업데이트하는 것을 가능하게 하며, 이는 사운드 이벤트 분류 모델이 사운드 클래스 내의 더 넓은 범위의 사운드들을 검출할 수 있게 한다. 드리프트는 사용자 디바이스가 직면했지만 사용자 디바이스에 의해 인식되지 않았던 사운드들에 대응할 수 있기 때문에, 사용자 디바이스가 사운드 클래스들의 이러한 직면된 변동들을 수용하도록 사운드 이벤트 분류 모델을 업데이트하는 것은 특정 사용자 디바이스가 일반적으로 직면한 사운드 클래스들의 특정 변동들을 더 정확하게 식별하는 것을 가능하게 한다.

[0012] [0027] 특정 양상에 따르면, 사운드 이벤트 분류 모델이 (사운드의 오디오 데이터 샘플들에 기반하여) 사운드의 사운드 클래스를 인식하지 않은 것으로 결정될 때, 사운드 이벤트 분류 모델이 사운드와 연관된 타입의 사운드 클래스들을 인식하지 못하기 때문에 또는 드리프트로 인해 사운드가 인식되지 않았는지 여부가 결정된다. 예컨대, 오디오 데이터 샘플들과 별개인 정보, 이를테면, 타임스탬프, 로케이션 데이터, 이미지 데이터, 비디오 데이터, 사용자 입력 데이터, 설정 데이터, 다른 센서 데이터 등이 오디오 데이터 샘플들과 연관된 사운드 환경(또는 오디오 장면)을 표시하는 장면 데이터를 결정하기 위해 사용된다. 장면 데이터는, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지(예컨대, 오디오 장면 내의 사운드 이벤트들을 인식하도록 트레이닝되는지) 여부를 결정하는 데 사용된다. 사운드 이벤트 분류 모델이 오디오 장면에 대응하는 경우, 오디오 데이터 샘플들은 모델 업데이트 데이터로서 저장되고, 드리프트 데이터인 것으로 표시된다. 일부 양상들에서, 사운드 이벤트 분류 모델이 오디오 장면에 대응하지 않으면, 오디오 데이터 샘플들은 미지로서 폐기되거나 또는 모델 업데이트 데이터로서 저장되고, 미지 사운드 클래스(예컨대, 미지 데이터)와 연관되는 것으로 표시된다.

[0013] [0028] 주기적으로 또는 가끔(예컨대, 사용자에 의해 개시될 때 또는 업데이트 조건이 충족될 때), 사운드 이벤트 분류 모델은 모델 업데이트 데이터를 사용하여 업데이트된다. 예컨대, 드리프트 데이터를 처리하기 위해, 사운드 이벤트 분류기는 역전파 또는 다른 유사한 머신-러닝 기법들을 사용하여 트레이닝될 수 있다(예컨대, 이미 트레이닝된 사운드 이벤트 분류기로 시작함으로써 추가로 트레이닝될 수 있음). 이 예에서, 드리프트 데이터는 사운드 이벤트 분류 모델에 의해 이미 인식된 사운드 클래스의 라벨과 연관되고, 드리프트 데이터 및 대응하는 라벨은 라벨링된 트레이닝 데이터로서 사용된다. 드리프트 데이터를 사용하여 사운드 이벤트 분류 모델을 업데이트하는 것은, 사운드 이벤트 분류 모델을 트레이닝시키기 위해 원래 사용된 트레이닝 데이터로부터 취해진 예들과 같은, 라벨링된 트레이닝 데이터에 사운드 클래스의 다른 예들을 부가함으로써 증강될 수 있다. 일부 양상들에서, 디바이스는 드리프트 데이터가 이용 가능할 때 하나 이상의 사운드 이벤트 분류 모델들을 자동으로(예컨대, 사용자 입력 없이) 업데이트한다. 따라서, 사운드 이벤트 분류 시스템은 사운드 이벤트 분류 모델을 처음부터 트레이닝하는 데 사용될 상당히 더 적은 컴퓨팅 자원들을 사용하여 사운드 클래스 내의 드리프트를 처리하도록 자동으로 구성될 수 있다.

[0014] [0029] 미지의 데이터를 처리하기 위해, 사운드 이벤트 분류 모델은 더 복잡한 전이 러닝 기법들을 사용하여 트레이닝될 수 있다. 예컨대, 미지의 데이터가 이용 가능할 때, 사용자가 사운드 이벤트 분류 모델을 업데이트하기를 원하는지 여부를 표시하도록 사용자가 질의 받을 수 있다. 미지의 데이터를 표현하는 오디오는 사용자에게 플레이 아웃될 수 있고, 사용자는 미지의 데이터가 사운드 이벤트 분류 모델을 업데이트하지 않고 폐기될 것임을 표시할 수 있거나, (예컨대, 미지의 데이터를 드리프트 데이터로서 재분류하기 위해) 미지의 데이터가 알려진 사운드 클래스에 대응한다고 표시할 수 있거나, 또는 미지의 데이터에 새로운 사운드 클래스 라벨을 할당할 수 있다. 사용자가 미지의 데이터를 드리프트 데이터로서 재분류하면, 앞서 설명한 바와 같이, 드리프트 데이터를 처리하기 위해 사운드 이벤트 분류 모델을 업데이트하는 데 사용되는 머신-러닝 기법(들)이 개시된다.

[0015] [0030] 사용자가 미지의 데이터에 새로운 사운드 클래스 라벨을 할당하면, 라벨 및 미지의 데이터는 업데이트된 사운드 이벤트 분류 모델을 생성하기 위해 라벨링된 트레이닝 데이터로서 사용된다. 특정 양상에 따르면, 사운드 이벤트 분류 모델을 업데이트하는 데 사용되는 전이 러닝 기법은 새로운 사운드 클래스와 연관된 출력 노드를 포함하는 사운드 이벤트 분류기 모델의 사본을 생성하는 것을 포함한다. 사운드 이벤트 분류기 모델의 카피는 증분 모델로 지칭된다. 전이 러닝 기법은 또한 사운드 이벤트 분류 모델 및 증분 모델을 하나 이상의 어댑터 네트워크들에 접속하는 것을 포함한다. 어댑터 네트워크(들)는 사운드 이벤트 분류 모델의 출력 및 증분 모델의 출력 둘 모두에 기반하는 병합된 출력의 생성을 가능하게 한다. 미지의 데이터를 포함하는 오디오 데이터 샘플들, 및 알려진 사운드 클래스들(예컨대, 사운드 이벤트 분류기가 사전에 트레이닝되어 인식하는 사운드 클래스들)에 대응하는 하나 이상의 오디오 데이터 샘플들은 사운드 이벤트 분류 모델 및 증분 모델에 제공되어, 병합된 출력을 생성한다. 병합된 출력은 사운드 이벤트 분류 모델, 증분 모델 및 하나 이상의 어댑터 네트워크들에 의한 분석에 기반하여 오디오 데이터 샘플들에 할당된 사운드 클래스를 표시한다. 트레이닝 동안, 병합된 출력은 증분 모델 및 어댑터 네트워크(들)의 링크 가중치들을 업데이트하는 데 사용된다. 트레이닝이 완료될 때, 증분 모델이 충분히 정확하다면, 사운드 이벤트 분류기는 폐기될 수 있다. 증분 모델 단독이 충분히 정확하지 않으면, 사운드 이벤트 분류 모델, 증분 모델 및 어댑터 네트워크(들)는 함께 유지되어 단일 업데이트된 사운드 이벤트 분류 모델로서 사용된다. 따라서, 본 명세서에 개시된 기법들은, 뉴럴 네트워크를 처음부터 트레이닝하는 것보다 (메모리 자원들, 프로세서 시간 및 전력 측면에서) 덜 자원 집약적인 방식으로 사운드 이벤트 분류 모델들의 맞춤화 및 업데이트를 가능하게 한다. 부가적으로, 일부 양상들에서, 개시된 기법들은 드리프트를 처리하기 위한 사운드 이벤트 분류 모델들의 자동 업데이트를 가능하게 한다.

[0016] [0031] 개시된 시스템들 및 방법들은 데이터세트 드리프트를 검출하고, (예컨대, 이용 가능한 멀티-모달 입력들을 활용함으로써) 드리프트 데이터를 대응하는 클래스와 연관시키고, 감독을 거의 또는 전혀 하지 않으면서 그리고 새로운 SEC 모델을 처음부터 트레이닝시키지 않으면서 드리프트 데이터를 사용하여 SEC 모델을 세밀화/미세-조정할 수 있는 상황-인식 시스템을 제공한다. 일부 양상들에서, SEC 모델을 세밀화/미세-조정하기 전에, SEC 모델은 특정 사운드 클래스의 다수의 변형들을 인식하도록 트레이닝되고, SEC 모델을 세밀화/미세-조정은 SEC 모델이 특정 사운드 클래스의 부가적인 변형을 인식할 수 있게 하기 위해 SEC 모델을 수정한다.

[0017] [0032] 일부 양상들에서, 개시된 시스템들 및 방법들은 테스트 동안 데이터세트 드리프트를 겪는 애플리케이션들에 사용될 수 있다. 예컨대, 시스템들 및 방법들은 데이터세트 드리프트를 검출하고 이전에 학습된 사운드 클래스들을 처음부터 재트레이닝하지 않고 SEC 모델을 개선할 수 있다. 일부 양상들에서, 개시된 시스템들 및 방법들은, SEC 모델을 처음부터 재트레이닝할 필요 없이, SEC 모델을 초기에 트레이닝하는 데 사용된 모든 트레이닝 데이터에 액세스할 필요 없이, 그리고 SEC 모델이 초기에 트레이닝되어 인식할 사운드 클래스들에 관련하여

어떠한 성능 저하도 도입하지 않으면서, 개시된 시스템들 및 방법들은 기존의 SEC 모델(예컨대, 특정 사운드 클래스들에 대해 이미 트레이닝된 SEC 모델)에 새로운 사운드 클래스들을 추가하는 데 사용될 수 있다.

[0018] [0033] 일부 양상들에서, 개시된 시스템들 및 방법들은, 낮은 풋프린트 제약에서의 연속적인 학습 능력이 바람직한 애플리케이션들에서 사용될 수 있다. 일부 구현들에서, 개시된 시스템들 및 방법들은 다양한 범위의 애플리케이션들(예컨대, 다양한 사운드 환경들)에 대한 다양한 검출 모델들(예컨대, SEC 모델들)의 데이터베이스에 액세스할 수 있다. 이러한 구현들에서, 동작 동안, 사운드 환경에 기반하여 SEC 모델이 선택될 수 있고, SEC 모델은 소스 모델로서 로딩 및 활용될 수 있다.

[0019] [0034] 도면들을 참조하여 본 개시의 특정 양상들이 아래에서 설명된다. 설명에서, 공통 특징들은 공통 참조 번호들로 지정된다. 본 명세서에서 사용되는 바와 같이, 다양한 용어는 오직 특정 구현들을 설명하기 위해 사용되며, 구현들의 제한으로 의도되지 않는다. 예를 들어, 본 명세서에서 사용되는 바와 같이, 단수형 형태들은, 문맥상 명확하게 달리 표시되지 않으면, 복수형 형태들을 또한 포함하도록 의도된다. 추가로, 본원에 설명된 일부 특징들은 일부 구현들에서는 단수이고, 다른 구현들에서는 복수이다. 예시하자면, 도 1은 하나 이상의 센서들을 포함하는 디바이스(100)(도 1의 "센서(들)(134))를 도시하는데, 이는 일부 구현들에서 디바이스(100)가 단일 센서(134)를 포함하고 다른 구현들에서 디바이스(100)가 다수의 센서들(134)을 포함함을 표시한다. 본 명세서에서의 참조의 용이함을 위해, 이러한 특징들은 일반적으로 "하나 이상의" 특징들로서 도입되고, 후속하여, 특징들 중 다수와 관련된 양상들이 설명되고 있지 않으면, 단수 또는 선택적인 복수(일반적으로 "(들)"로 끝나는 용어들로 표시됨)로 지칭된다.

[0020] [0035] "포함하다(comprise)" 및 "포함하는(comprising)"이라는 용어들은 본 명세서에서 "포함한다(include)", 또는 "포함하는(including)"과 상호 교환 가능하게 사용된다. 추가로, "~여기서"라는 용어는 "~에서"와 상호 교환 가능하게 사용된다. 본 명세서에서 사용되는 바와 같이, “예시적인” 은 예, 구현 및/또는 양상을 표시하고, 제한적인 것으로 또는 선호도 또는 선호되는 구현을 표시하는 것으로 해석되어서는 안 된다. 본 명세서에서 사용되는 바와 같이, 엘리먼트, 예를 들어, 구조, 컴포넌트, 동작 등을 수정하기 위해 사용되는 서수적 용어(예를 들어, "제1", "제2", "제3" 등)는 그 자체로 엘리먼트의 다른 엘리먼트에 대한 어떠한 우선순위 또는 순서를 표시하는 것이 아니라, 오히려 엘리먼트를 동일 이름을 갖는(그러나 서수적 용어를 사용한) 다른 엘리먼트로부터 단순히 구별한다. 본 명세서에서 사용되는 바와 같이, "세트"라는 용어는 특정 엘리먼트 중 하나 이상을 지칭하고, "복수"라는 용어는 다수의(예를 들어, 둘 이상의) 특정 엘리먼트를 지칭한다.

[0021] [0036] 본원에서 사용되는 바와 같이, "결합된"은 "통신 가능하게 결합된", "전기적으로 결합된" 또는 "물리적으로 결합된"을 포함할 수 있고, 또한(또는 대안적으로) 이들의 임의의 조합들을 포함할 수 있다. 2개의 디바이스들(또는 컴포넌트들)은 하나 이상의 다른 디바이스들, 컴포넌트들, 와이어들, 버스들, 네트워크들(예컨대, 유선 네트워크, 무선 네트워크, 또는 이들의 조합) 등을 통해 직접적으로 또는 간접적으로 결합될(예컨대, 통신 가능하게 결합, 전기적으로 결합 또는 물리적으로 결합) 수 있다. 예시적인 비-제한적인 예들로서, 전기적으로 커플링되는 2개의 디바이스들(또는 컴포넌트들)은 동일한 디바이스에 또는 상이한 디바이스들에 포함될 수 있고, 전자기기(electronics), 하나 이상의 커넥터들, 또는 유도성 커플링을 통해 연결될 수 있다. 일부 구현들에서, 이를테면, 전기 통신으로 통신 가능하게 결합된 2개의 디바이스들(또는 컴포넌트들)은, 직접적으로 또는 간접적으로, 이를테면, 하나 이상의 와이어들, 버스들, 네트워크들 등을 통해, 전기 신호들(디지털 신호들 또는 아날로그 신호들)을 전송 및 수신할 수 있다. 본원에서 사용되는 바와 같이, "직접적으로 결합된"은 개재 컴포넌트들 없이 결합된(예컨대, 통신 가능하게 결합된, 전기적으로 결합된, 또는 물리적으로 결합된) 2개의 디바이스들을 지칭한다.

[0022] [0037] 본 개시내용에서, "결정", "계산", "추정", "시프팅", "조정" 등과 같은 용어들은 하나 이상의 동작들이 어떻게 수행되는지를 설명하기 위해 사용될 수 있다. 이러한 용어들은 제한적인 것으로 해석되지 않아야 하며, 다른 기법들이 유사한 동작들을 수행하기 위해 활용될 수 있다는 것이 주목되어야 한다. 추가로, 본 명세서에서 언급되는 바와 같이, "생성하는", "계산하는", "추정하는", "사용하는", "선택하는", "액세스하는" 및 "결정하는"은 상호 교환 가능하게 사용될 수 있다. 예컨대, 파라미터(또는 신호)를 "생성하는", "계산하는", "추정하는" 또는 "결정하는" 것은 파라미터(또는 신호)를 능동적으로 생성, 추정, 계산 또는 결정하는 것을 지칭할 수 있거나, 이를테면, 다른 컴포넌트 또는 디바이스에 의해 이미 생성된 파라미터(또는 신호)를 선택하거나 또는 액세스하는 것을 지칭할 수 있다.

[0023] [0038] 도 1은 오디오 데이터 샘플들(110)에 대한 응답으로 사운드 식별 데이터를 생성하도록 구성되고 사운드 이벤트 분류 모델을 업데이트하도록 구성된 디바이스(100)의 예의 블록도이다. 도 1의 디바이스(100)는 음향 환

경 내에서 검출된 사운드(102)에 기반하여 오디오 신호들(106)을 생성하도록 구성된 하나 이상의 마이크로폰들(104)을 포함한다. 마이크로폰(들)(104)은 오디오 신호들(106)에 기반하여 오디오 데이터 샘플들(110)을 생성하는 특징 추출기(108)에 결합된다. 예컨대, 오디오 데이터 샘플들(110)은 데이터 엘리먼트들의 어레이 또는 매트릭스를 포함할 수 있으며, 각각의 데이터 엘리먼트는 오디오 신호들(106)에서 검출된 특징에 대응한다. 특정 예로서, 오디오 데이터 샘플들(110)은 오디오 신호들(106) 중 1초로부터 추출된 멜(Mel) 스펙트럼 특징들에 대응할 수 있다. 이러한 예에서, 오디오 데이터 샘플들(110)은 특징 값들의 128x128 엘리먼트 행렬을 포함할 수 있다. 다른 예들에서, 다른 오디오 데이터 샘플 구성들 또는 크기들이 사용될 수 있다.

[0024] [0039] 오디오 데이터 샘플들(110)은 SEC(sound event classification) 엔진(120)에 제공된다. SEC 엔진(120)은 SEC 모델(112)과 같은 하나 이상의 SEC 모델들에 기반하여 추론 동작들을 수행하도록 구성된다. "추론 동작들"은, 오디오 데이터 샘플들(110)의 사운드 클래스가 SEC 모델(112)에 의해 인식되는 경우, 오디오 데이터 샘플들(110)을 사운드 클래스에 할당하는 것을 지칭한다. 예컨대, SEC 엔진(120)은 머신-러닝 런타임 환경을 구현하는 소프트웨어, 이를테면, 미국 캘리포니아주 샌디에이고의 QUALCOMM Technologies, Inc.로부터 입수 가능한 Qualcomm Neural Processing SDK를 포함하거나 그에 대응할 수 있다. 특정 양상에서, SEC 모델(112)은 SEC 엔진(120)에 이용 가능한 복수의 SEC 모델들(예컨대, 이용 가능한 SEC 모델들(114)) 중 하나이다.

[0025] [0040] 특정 예에서, 이용 가능한 SEC 모델들(114) 각각은 사운드 이벤트 분류기로서 트레이닝된 뉴럴 네트워크를 포함하거나 그에 대응한다. 예시하자면, SEC 모델(112)(뿐만 아니라 다른 이용 가능한 SEC 모델들(114) 각각)은 입력 계층, 하나 이상의 은닉 계층들 및 출력 계층을 포함할 수 있다. 이 예에서, 입력 계층은 특징 추출기(108)에 의해 생성된 오디오 데이터 샘플들(110)의 값들의 어레이 또는 행렬에 대응하도록 구성된다. 예시하자면, 오디오 데이터 샘플들(110)이 15개의 데이터 엘리먼트들을 포함하면, 입력 계층은 15개의 노드들(예컨대, 데이터 엘리먼트 당 하나)을 포함할 수 있다. 출력 계층은, SEC 모델(112)이 트레이닝되어 인식할 사운드 클래스들에 대응하도록 구성된다. 출력 층의 특정 어레인지먼트는 출력으로서 제공될 정보에 따라 변할 수 있다. 일례로, SEC 모델(112)은 사운드 클래스당 하나의 비트를 포함하는 어레이를 출력하도록 트레이닝될 수 있으며, 여기서 출력 계층은 출력 어레이의 비트들 중 하나를 제외한 모든 비트가 0의 값을 갖도록 "원 핫 인코딩"을 수행하며, 검출된 사운드 클래스에 대응하는 비트는 1의 값을 갖는다. 예컨대, 각각의 사운드 클래스에 대한 신뢰도 메트릭의 값을 표시하기 위해 다른 출력 방식들이 사용될 수 있으며, 여기서 신뢰도 메트릭의 값은 오디오 데이터 샘플들(110)이 개개의 사운드 클래스에 대응한다는 확률 추정치를 표시한다. 예시하자면, SEC 모델(112)이 4개의 사운드 클래스들을 인식하도록 트레이닝되면, SEC 모델(112)은 4개의 값들(사운드 클래스당 하나)을 포함하는 출력 데이터를 생성할 수 있고, 각각의 값은 오디오 데이터 샘플들(110)이 개개의 사운드 클래스에 대응한다는 확률 추정을 표시할 수 있다.

[0026] [0041] 은닉 층들 각각은 복수의 노드들을 포함하고, 각각의 노드는 동일한 층 또는 상이한 층의 다른 노드들과 (링크를 통해) 상호연결된다. 노드의 각각의 입력 링크는 링크 가중치와 연관된다. 동작 동안, 노드는 자신이 링크된 다른 노드들로부터 입력 값들을 수신하고, 결합된 값을 결정하기 위해 대응하는 링크 가중치들에 기반하여 입력 값들을 가중하고, 결합된 값을 활성화 함수에 공급하여 노드의 출력 값을 생성한다. 출력 값은 노드의 출력 링크들을 통해 하나 이상의 다른 노드들에 제공된다. 노드들은 또한 결합된 값을 생성하는 데 사용되는 바이어스 값들을 포함할 수 있다. 노드들은 다양한 어레인지먼트(arrangement)들로 링크될 수 있고, 특정 데이터의 프로세싱을 용이하게 하기 위해 다양한 다른 특징들(예컨대, 이전 값들의 메모리)을 포함할 수 있다. 오디오 데이터 샘플들의 경우, CNN(convolutional neural network)들이 사용될 수 있다. 예시하자면, SEC 모델들(112) 중 하나 이상은 3개의 링크된 CNN들을 포함할 수 있고, 각각의 CNN은 2차원(2D) 콘볼루션 계층, 맥스풀링 계층 및 배치 정규화 계층을 포함할 수 있다. 다른 구현들에서, 은닉 계층들은 상이한 수의 CNN들 또는 다른 계층들을 포함한다. 뉴럴 네트워크를 트레이닝시키는 것은 뉴럴 네트워크의 출력 에러를 감소시키기 위해 링크 가중치들을 수정하는 것을 포함한다.

[0027] [0042] 동작 동안, SEC 엔진(120)은 오디오 데이터 샘플들(110)을 입력으로서 단일 SEC 모델(예컨대, SEC 모델(112)), 다수의 선택된 SEC 모델들(예컨대, SEC 모델(112) 및 이용 가능한 SEC 모델들(114)의 제K SEC 모델(118)), 또는 SEC 모델들(예컨대, SEC 모델(112), 제1 SEC 모델(116), 제K SEC 모델(118), 및 이용 가능한 SEC 모델들(114) 중 임의의 다른 SEC 모델들) 각각에 제공할 수 있다. 예컨대, SEC 엔진(120)(또는 디바이스(100)의 다른 컴포넌트)은 예컨대, 사용자 입력, 디바이스(100)와 연관된 디바이스 설정들, 센서 데이터, 오디오 데이터 샘플들(110)이 수신되는 시간, 또는 다른 팩터들에 기반하여, 이용 가능한 SEC 모델들(114) 중에서 SEC 모델(112)을 선택할 수 있다. 이러한 예에서, SEC 엔진(120)은 SEC 모델(112)만을 사용하도록 선택할 수 있거나 이용 가능한 SEC 모델들(114) 중 2개 이상을 사용하도록 선택할 수 있다. 예시하자면, 디바이스 설정들은 SEC 모

델(112) 및 제1 SEC 모델(116)이 특정 시간 프레임 동안 사용될 것임을 표시할 수 있다. 다른 예에서, SEC 엔진(120)은 각각으로부터의 출력을 생성하기 위해 오디오 데이터 샘플들(110)을 이용 가능한 SEC 모델들(114) 각각에 (예컨대, 순차적으로 또는 병렬로) 제공할 수 있다. 특정 양상에서, SEC 모델들은 상이한 사운드 클래스들을 인식하도록, 상이한 음향 환경들에서 동일한 사운드 클래스들을 인식하도록, 또는 둘 모두를 위해 트레이닝된다. 예컨대, SEC 모델(112)은 제1 세트의 사운드 클래스들을 인식하도록 구성될 수 있고, 제1 SEC 모델(116)은 제2 세트의 사운드 클래스들을 인식하도록 구성될 수 있으며, 여기서 제1 세트의 사운드 클래스들은 제2 세트의 사운드 클래스들과 상이하다.

[0028] [0043] 특정 양상에서, SEC 엔진(120)은 SEC 모델(112)의 출력에 기반하여, SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식하였는지 여부를 결정한다. SEC 엔진(120)이 오디오 데이터 샘플들(110)을 다수의 SEC 모델들에 제공하면, SEC 엔진(120)은 SEC 모델들 각각의 출력에 기반하여, SEC 모델들 중 임의의 것이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식하였는지 여부를 결정할 수 있다. SEC 모델(112)(또는 이용 가능한 SEC 모델들(114) 중 다른 모델)이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식했다면, SEC 엔진(120)은 오디오 데이터 샘플들(110)의 사운드 클래스(122)를 표시하는 출력(124)을 생성한다. 예컨대, 출력(124)은 사운드(102)와 연관된 사운드 클래스(122)의 검출을 사용자에게 통지하기 위해 디스플레이에 전송될 수 있거나, 또는 다른 디바이스 또는 디바이스(100)의 다른 컴포넌트에 전송되어 (예컨대, 도어 폐쇄의 사운드를 인식하는 것에 대한 응답으로 조명들을 활성화시키기 위한 커맨드를 전송하기 위한) 액션을 트리거하도록 사용될 수 있다.

[0029] [0044] SEC 모델(112)(및 오디오 데이터 샘플들(110)이 제공된 이용 가능한 SEC 모델들(114) 중 다른 모델들)이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식하지 못했다고 SEC 엔진(120)이 결정하는 경우, SEC 엔진(120)은 트리거 신호(126)를 드리프트 검출기(128)에 제공한다. 예컨대, SEC 엔진(120)은 디바이스(100)의 메모리에서 트리거 플래그를 세팅할 수 있다. 일부 구현들에서, SEC 엔진(120)은 또한 다른 데이터를 드리프트 검출기(128)에 제공할 수 있다. 예시하자면, SEC 모델(112)이 트레이닝되어 인식할 각각의 사운드 클래스에 대한 신뢰도 메트릭의 값을 SEC 모델(112)이 생성하면, 신뢰도 메트릭의 값들 중 하나 이상이 드리프트 검출기(128)에 제공될 수 있다. 예컨대, SEC 모델(112)이 3개의 사운드 클래스들을 인식하도록 트레이닝되면, SEC 엔진(120)은 SEC 모델(112)에 의해 출력된 3개의 신뢰도 값들(3개의 사운드 클래스들 각각에 대해 하나) 중에서 가장 높은 신뢰도 값을 드리프트 검출기(128)에 제공할 수 있다.

[0030] [0045] 특정 양상에서, SEC 엔진(120)은 신뢰도 메트릭의 값에 기반하여 SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식했는지 여부를 결정한다. 이러한 특정 양상에서, 특정 사운드 클래스에 대한 신뢰도 메트릭의 값은 오디오 데이터 샘플들(110)이 특정 사운드 클래스와 연관될 확률을 표시한다. 예시하자면, SEC 모델(112)이 4개의 사운드 클래스들을 인식하도록 트레이닝되면, SEC 모델(112)은 각각의 사운드 클래스에 대해 하나씩, 신뢰도 메트릭의 4개의 값들을 포함하는 어레이를 출력으로서 생성할 수 있다. 일부 구현들에서, SEC 엔진(120)은, 사운드 클래스(122)에 대한 신뢰도 메트릭의 값이 검출 임계치보다 크면, SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스(122)를 인식하였다고 결정한다. 예컨대, SEC 엔진(120)은, 사운드 클래스(122)에 대한 신뢰도 메트릭의 값이 0.90(예컨대, 90% 신뢰도), 0.95(예컨대, 95% 신뢰도), 또는 검출 임계치의 어떤 다른 값보다 크면, SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스(122)를 인식하였다고 결정한다. 일부 구현들에서, SEC 모델(112)이 트레이닝되어 인식할 각각의 사운드 클래스에 대한 신뢰도 메트릭의 값이 검출 임계치보다 작으면, SEC 엔진(120)은 SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스를 인식하지 않았다고 결정한다. 예컨대, SEC 엔진(120)은, 신뢰도 메트릭의 각각의 값이 0.90(예컨대, 90% 신뢰도), 0.95(예컨대, 95% 신뢰도), 또는 검출 임계치의 어떤 다른 값보다 작으면, SEC 모델(112)이 오디오 데이터 샘플들(110)의 사운드 클래스(122)를 인식하지 않았다고 결정한다.

[0031] [0046] 드리프트 검출기(128)는 오디오 데이터 샘플들(110)의 사운드 클래스를 인식할 수 없었던 SEC 모델(112)이 오디오 데이터 샘플들(110)과 연관된 오디오 장면(142)에 대응하는지 여부를 결정하도록 구성된다. 도 1에 예시된 예에서, 장면 검출기(140)는 장면 데이터(138)를 수신하고 장면 데이터(138)를 사용하여 오디오 데이터 샘플들(110)과 연관된 오디오 장면(142)을 결정하도록 구성된다. 특정 양상에서, 장면 데이터(138)는, 디바이스(100)와 연관된 하나 이상의 디바이스 설정들을 나타내는 설정 데이터(130), 클록(132)의 출력, 하나 이상의 센서들(134)로부터의 센서 데이터, 입력 디바이스(136)를 통해 수신된 입력, 또는 이들의 조합에 기반하여 생성된다. 일부 양상들에서, 장면 검출기(140)는, SEC 엔진(120)이 SEC 모델(112)을 선택하기 위해 사용하는 것과 상이한 정보를 사용하여 오디오 장면(142)을 결정한다. 예시하자면, SEC 엔진(120)이 하루 중의 시각에 기반하여 SEC 모델(112)을 선택하면, 장면 검출기(140)는 오디오 장면(142)을 결정하기 위해 센서(들)(134)의 포지션 센

서로부터의 포지션 센서 데이터를 사용할 수 있다. 일부 양상들에서, 장면 검출기(140)는, SEC 엔진(120)이 SEC 모델(112)을 선택하기 위해 사용하는 것과 동일한 정보 중 적어도 일부 및 부가적인 정보를 사용한다. 예시하자면, SEC 엔진(120)이 하루 중의 시각 및 세팅 데이터(130)에 기반하여 SEC 모델(112)을 선택하면, 장면 검출기(140)는 오디오 장면(142)을 결정하기 위해 포지션 센서 데이터 및 세팅 데이터(130)를 사용할 수 있다. 따라서, 장면 검출기(140)는 SEC 모델(112)을 선택하기 위해 SEC 엔진(120)에 의해 사용되는 것과 상이한 오디오 장면 검출 모드를 사용한다.

[0032] [0047] 특정 구현에서, 장면 검출기(140)는 장면 데이터(138)에 기반하여 오디오 장면(142)을 결정하도록 트레이닝된 뉴럴 네트워크이다. 다른 구현들에서, 장면 검출기(140)는 상이한 머신-러닝 기법을 사용하여 트레이닝된 분류기이다. 예컨대, 장면 검출기(140)는 결정 트리, 랜덤 포리스트, 지원 벡터 머신, 또는 장면 데이터(138)에 기반하여 오디오 장면(142)을 표시하는 출력을 생성하도록 트레이닝된 다른 분류기를 포함하거나 그에 대응할 수 있다. 또 다른 구현들에서, 장면 검출기(140)는 장면 데이터(138)에 기반하여 오디오 장면(142)을 결정하기 위해 휴리스틱(heuristics)을 사용한다. 또 다른 구현들에서, 장면 검출기(140)는 장면 데이터(138)에 기반하여 오디오 장면(142)을 결정하기 위해 인공 지능과 휴리스틱의 조합을 사용한다. 예컨대, 장면 데이터(138)는 이미지 데이터, 비디오 데이터, 또는 둘 모두를 포함할 수 있고, 장면 검출기(140)는 특정 객체들, 모션들, 배경들, 또는 다른 이미지나 비디오 정보를 검출하기 위해 머신-러닝 기법을 사용하여 트레이닝되는 이미지 인식 모델을 포함할 수 있다. 이러한 예에서, 이미지 인식 모델의 출력은 오디오 장면(142)을 결정하기 위해 하나 이상의 휴리스틱을 통해 평가될 수 있다.

[0033] [0048] 드리프트 검출기(128)는 장면 검출기(140)에 의해 표시된 오디오 장면(142)을 SEC 모델(112)을 설명하는 정보와 비교하여, SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)과 연관되는지 여부를 결정한다. SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)과 연관된다고 드리프트 검출기(128)가 결정하면, 드리프트 검출기(128)는 드리프트 데이터(144)가 모델 업데이트 데이터(148)로서 저장되게 한다. 특정 구현에서, 드리프트 데이터(144)는 오디오 데이터 샘플들(110) 및 라벨 - 라벨은 SEC 모델(112)을 식별함 - 을 포함하거나, 오디오 데이터 샘플들(110)과 연관된 사운드 클래스를 표시하거나, 또는 둘 모두이다. 드리프트 데이터(144)가 오디오 데이터 샘플들(110)과 연관된 사운드 클래스를 표시하면, 사운드 클래스는 SEC 모델(112)에 의해 생성된 신뢰도 메트릭의 최고 값에 기반하여 선택될 수 있다. 예시적인 예로서, SEC 엔진(120)이 0.90의 검출 임계치를 사용하고, SEC 모델(112)에 의해 출력된 신뢰도 메트릭의 최고 값이 특정 사운드 클래스에 대해 0.85인 경우, SEC 엔진(120)은 오디오 데이터 샘플들(110)의 사운드 클래스가 인식되지 않고 트리거 신호(126)를 드리프트 검출기(128)에 전송한다고 결정한다. 이 예에서, 드리프트 검출기(128)가 SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)에 대응한다고 결정하면, 드리프트 검출기(128)는 그 오디오 데이터 샘플들(110)을 특정 사운드 클래스와 연관된 드리프트 데이터(144)로서 저장한다. 특정 양상에서, SEC 모델들(114)과 연관된 메타데이터는 각각의 SEC 모델(114)과 연관된 오디오 장면 또는 오디오 장면들을 특정하는 정보를 포함한다. 예컨대, SEC 모델(112)은 사용자의 집에서 사운드 이벤트들을 검출하도록 구성될 수 있으며, 이 경우에, SEC 모델(112)과 연관된 메타데이터는 SEC 모델(112)이 "홈" 오디오 장면과 연관된다는 것을 표시할 수 있다. 이러한 예에서, (예컨대, 포지션 정보, 사용자 입력, 홈 무선 네트워크 신호의 검출, 홈 위치들을 나타내는 이미지 또는 비디오 데이터 등에 기반하여) 디바이스(100)가 홈 로케이션에 있다고 오디오 장면(142)이 표시하면, 드리프트 검출기(128)는 SEC 모델(112)이 오디오 장면(142)에 대응한다고 결정한다.

[0034] [0049] 일부 구현들에서, 드리프트 검출기(128)는 또한 일부 오디오 데이터 샘플들(110)이 모델 업데이트 데이터(148)로서 저장되고 미지의 데이터(146)로 지정되게 한다. 제1 예로서, SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)에 대응하지 않는다고 드리프트 검출기(128)가 결정하면, 드리프트 검출기(128)는 미지의 데이터(146)를 저장할 수 있다. 제2 예로서, 드리프트 검출기(128)는, SEC 모델(112)에 의해 출력된 신뢰도 메트릭의 값이 드리프트 임계치를 충족하는 것을 실패하면, 미지의 데이터(146)를 저장할 수 있다. 이러한 예에서, 드리프트 임계치는 SEC 엔진(120)에 의해 사용되는 검출 임계치보다 작다. 예컨대, SEC 엔진(120)이 0.95의 검출 임계치를 사용하면, 드리프트 임계치는 0.80의 값, 0.75의 값, 또는 0.95 미만의 어떤 다른 값을 가질 수 있다. 이 예에서, 오디오 데이터 샘플들(110)에 대한 신뢰도 메트릭의 최고 값이 드리프트 임계치보다 작으면, 드리프트 검출기(128)는 오디오 데이터 샘플들(110)이, SEC 모델(112)이 인식을 위해 트레이닝되지 않은 사운드 클래스에 속한다고 결정하고 오디오 데이터 샘플들(110)을 미지의 데이터(146)로서 지정한다. 특정 양상에서, SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)에 대응한다고 드리프트 검출기(128)가 결정하면, 드리프트 검출기(128)는 미지의 데이터(146)만을 저장한다. 다른 특정 양상에서, SEC 모델(112)이 오디오 데이터 샘플들(110)의 오디오 장면(142)에 대응한다고 드리프트 검출기(128)가 결정하는지 여부와

독립적으로, 드리프트 검출기(128)는 미지의 데이터(146)를 저장한다.

[0035] [0050] 모델 업데이트 데이터(148)가 저장된 후, 모델 업데이터(152)는 모델 업데이트 데이터(148)에 액세스하고 모델 업데이트 데이터(148)를 사용하여 이용 가능한 SEC 모델들(114) 중 하나(예컨대, SEC 모델(112))를 업데이트할 수 있다. 예컨대, 모델 업데이트 데이터(148)의 각각의 엔트리는 엔트리가 연관되는 SEC 모델을 표시하고, 모델 업데이터(152)는 대응하는 SEC 모델을 업데이트하기 위한 트레이닝 데이터로서 엔트리를 사용한다. 특정 양상에서, 모델 업데이터(152)는, 업데이트 기준이 충족될 때, 또는 모델 업데이트가 사용자 또는 다른 당사자(예컨대, 디바이스(100)의 벤더, SEC 엔진(120), SEC 모델들(114) 등)에 의해 개시될 때 SEC 모델을 업데이트한다. 업데이트 기준은, 모델 업데이트 데이터(148)에서 특정 수의 엔트리들이 이용 가능할 때, 특정 SEC 모델에 대한 특정 수의 엔트리들이 모델 업데이트 데이터(148)에서 이용 가능할 때, 특정 사운드 클래스에 대한 특정 수의 엔트리들이 모델 업데이트 데이터(148)에서 이용 가능할 때, 이전 업데이트 이후 특정 양의 시간이 경과했을 때, 다른 업데이트들이 발생할 때(예컨대, 디바이스(100)와 연관된 소프트웨어 업데이트가 발생할 때), 또는 다른 이벤트의 발생에 기반하여 충족될 수 있다.

[0036] [0051] 모델 업데이터(152)는 역전파 또는 유사한 머신-러닝 최적화 프로세스를 사용하여 SEC 모델(112)의 트레이닝을 업데이트하기 위해 라벨링된 트레이닝 데이터로서 드리프트 데이터(144)를 사용한다. 예컨대, 모델 업데이터(152)는 모델 업데이트 데이터(148)의 드리프트 데이터(144)로부터의 오디오 데이터 샘플들을 입력으로서 SEC 모델(112)에 제공하고, SEC 모델(112)의 출력 및 (드리프트 검출기(128)에 의해 저장된 드리프트 데이터(144)에 표시된 바와 같이) 오디오 데이터 샘플들과 연관된 라벨에 기반하여 (손실 함수로 또한 지칭되는) 에러 함수의 값을 결정하고, 변화도 하강 동작(또는 그 일부 변형) 또는 다른 기계-학습 최적화 프로세스를 사용하여 SEC 모델(112)에 대한 업데이트된 링크 가중치들을 결정한다.

[0037] [0052] 모델 업데이터(152)는 또한 업데이트 트레이닝 동안 SEC 모델(112)에 (드리프트 데이터(144)의 오디오 데이터 샘플들에 추가로) 다른 오디오 데이터 샘플들을 제공할 수 있다. 예컨대, 모델 업데이트 데이터(148)는 하나 이상의 알려진 오디오 데이터 샘플들(이를테면, SEC 모델(112)을 트레이닝시키기 위해 원래 사용된 오디오 데이터 샘플들의 서브세트)을 포함할 수 있고, 이는 업데이트 트레이닝이 SEC 모델(112)로 하여금 이전의 트레이닝을 망각하게 할 가능성들을 감소시킬 수 있다(여기서 "망각"은 SEC 모델(112)이 이전에 트레이닝되어 인식할 트레이닝된 사운드 클래스들을 검출하기 위한 신뢰성을 잃는 것을 지칭함). 드리프트 데이터(144)의 오디오 데이터 샘플들과 연관된 사운드 클래스가 드리프트 검출기(128)에 의해 표시되기 때문에, 드리프트를 처리하기 위한 업데이트 트레이닝은 (예컨대, 사용자 입력 없이) 자동으로 달성될 수 있다. 결과적으로, 디바이스(100)의 기능성(예컨대, 사운드 클래스들을 인식할 때의 정확도)은, 사용자 개입 없이 그리고 처음부터 새로운 SEC 모델을 생성하는 데 사용될 것보다 더 적은 컴퓨팅 자원들을 사용하여 시간이 지남에 따라 개선될 수 있다. 모델 업데이터(152)가 드리프트 데이터(144)에 기반하여 SEC 모델(112)을 업데이트하는 데 사용할 수 있는 전이 러닝 프로세스의 특정 예가 도 2를 참조하여 설명된다.

[0038] [0053] 일부 양상들에서, 모델 업데이터(152)는 또한 SEC 모델(112)의 트레이닝을 업데이트하기 위해 모델 업데이트 데이터(148)의 미지의 데이터(146)를 사용할 수 있다. 예컨대, 주기적으로 또는 가끔, 이를테면 업데이트 기준이 충족될 때, 모델 업데이터(152)는 사용자에게 모델 업데이트 데이터(148) 내의 미지의 데이터(146)의 엔트리들의 사운드 클래스를 라벨링하도록 요청하기 위해 사용자를 프롬프팅(prompt)할 수 있다. 사용자가 미지의 데이터(146)의 엔트리의 사운드 클래스를 라벨링하기로 선택하면, 디바이스(100)(또는 다른 디바이스)는 미지의 데이터(146)의 오디오 데이터 샘플들에 대응하는 사운드를 플레이아웃할 수 있다. 사용자는 오디오 데이터 샘플들의 사운드 클래스를 식별하는 하나 이상의 라벨들(150)을 (예컨대, 입력 디바이스(136)를 통해) 제공할 수 있다. 사용자에 의해 표시된 사운드 클래스가 SEC 모델(112)이 트레이닝되어 인식할 사운드 클래스이면, 미지의 데이터(146)는 사용자-특정 사운드 클래스 및 SEC 모델(112)과 연관된 드리프트 데이터(144)로서 재분류된다. 모델 업데이터(152)의 구성에 의존하여, 사용자에 의해 표시된 사운드 클래스가 SEC 모델(112)이 인식하도록 트레이닝되지 않은 사운드 클래스이면(예컨대, 새로운 사운드 클래스이면), 모델 업데이터(152)는 미지의 데이터(146)를 폐기하고, 새로운 또는 업데이트된 SEC 모델을 생성하기 위해 사용할 다른 디바이스에 미지의 데이터(146) 및 사용자-특정 사운드 클래스를 전송할 수 있거나, 또는 SEC 모델(112)을 업데이트하기 위해 미지의 데이터(146) 및 사용자-특정 사운드 클래스를 사용할 수 있다. 모델 업데이터(152)가 미지의 데이터(146) 및 사용자-특정 사운드 클래스에 기반하여 SEC 모델(112)을 업데이트하는 데 사용할 수 있는 전이 러닝 프로세스의 특정 예가 도 3을 참조하여 설명된다.

[0039] [0054] 모델 업데이터(152)에 의해 생성된 업데이트된 SEC 모델(154)은 이용 가능한 SEC 모델들(114)에 추가되어, 업데이트된 SEC 모델(154)이 생성된 후에 수신된 오디오 데이터 샘플들(110)을 평가하는 데 업데이트된 SEC

모델(154)을 이용 가능하게 한다. 따라서, 사운드들을 평가하는 데 사용될 수 있는 한 세트의 이용 가능한 SEC 모델들(114)은 동적이다. 예컨대, 이용 가능한 SEC 모델들(114) 중 하나 이상은 드리프트 데이터(144)를 처리하도록 자동으로 업데이트될 수 있다. 부가적으로, 이용 가능한 SEC 모델들(114) 중 하나 이상은, 새로운 SEC 모델을 처음부터 트레이닝시키는 것보다 더 적은 컴퓨팅 자원들(예컨대, 메모리, 프로세싱 시간, 및 전력)을 사용하는 전이 러닝 동작들을 사용하여 미지의 사운드 클래스들을 처리하도록 업데이트될 수 있다.

[0040] [0055] 도 2는 특정 예에 따라 드리프트를 처리하기 위해 SEC 모델(208)을 업데이트하는 양상들을 예시하는 도면이다. 도 2의 SEC 모델(208)은 드리프트 데이터(144)와 연관된 도 1의 이용 가능한 SEC 모델들(114) 중 특정 모델을 포함하거나 그에 대응한다. 예컨대, SEC 엔진(120)이 SEC 모델(112)의 출력에 대한 응답으로 트리거 신호(126)를 생성했다면, 드리프트 데이터(144)는 SEC 모델(112)과 연관되고, SEC 모델(208)은 SEC 모델(112)에 대응하거나 SEC 모델(112)을 포함한다. 다른 예로서, SEC 엔진(120)이 SEC 모델(118)의 출력에 대한 응답으로 트리거 신호(126)를 생성했다면, 드리프트 데이터(144)는 제K SEC 모델(118)과 연관되고, SEC 모델(208)은 제K SEC 모델(118)에 대응하거나 SEC 모델(118)을 포함한다.

[0041] [0056] 도 2에 예시된 예에서, 트레이닝 데이터(202)는 SEC 모델(208)을 업데이트하는 데 사용된다. 트레이닝 데이터(202)는 드리프트 데이터(144) 및 하나 이상의 라벨들(204)을 포함한다. 드리프트 데이터(144)의 각각의 엔트리는 오디오 데이터 샘플들(예컨대, 오디오 데이터 샘플들(206))을 포함하고, 라벨(들)(204)의 대응하는 라벨과 연관된다. 드리프트 데이터(144)의 엔트리의 오디오 데이터 샘플들은 SEC 모델(208)에 의해 인식되지 않은 사운드로부터 추출되거나 그에 기반하여 결정된 특징들을 표현하는 값들의 세트를 포함한다. 드리프트 데이터(144)의 엔트리에 대응하는 라벨(204)은 사운드가 속할 것으로 예상되는 사운드 클래스를 식별한다. 예로서, 드리프트 데이터(144)의 엔트리에 대응하는 라벨(204)은, SEC 모델(208)이 오디오 데이터 샘플들이 생성된 오디오 장면에 대응한다는 결정에 대한 응답으로, 도 1의 드리프트 검출기(128)에 의해 할당될 수 있다. 이러한 예에서, 드리프트 검출기(128)는 오디오 데이터 샘플들을, SEC 모델(208)의 출력에서 최고 신뢰도 메트릭 값과 연관된 사운드 클래스에 할당할 수 있다.

[0042] [0057] 도 2에서, 사운드에 대응하는 오디오 데이터 샘플들(206)이 SEC 모델(208)에 제공되고, SEC 모델(208)은 오디오 데이터 샘플들(206)이 할당되는 사운드 클래스를 나타내는 출력(210), 하나 이상의 신뢰도 메트릭 값들, 또는 둘 모두를 생성한다. 모델 업데이터(152)는 오디오 데이터 샘플들(206)에 대응하는 라벨(204) 및 출력(210)을 사용하여, SEC 모델(208)에 대한 업데이트된 링크 가중치들(212)을 결정한다. SEC 모델(208)은 업데이트된 링크 가중치들(212)에 기반하여 업데이트되고, 트레이닝 프로세스는 트레이닝 종료 조건이 충족될 때까지 되풀이하여 반복된다. 트레이닝 동안, 드리프트 데이터(144)의 엔트리들 각각은 SEC 모델(208)에 제공될 수 있다(예컨대, 반복 당 하나의 엔트리). 부가적으로, 일부 구현들에서, 다른 오디오 데이터 샘플들(예컨대, SEC 모델(208)을 트레이닝시키기 위해 이전에 사용된 오디오 데이터 샘플들)은 또한, SEC 모델(208)이 이전 트레이닝을 망각할 가능성을 감소시키기 위해 SEC 모델(208)에 제공될 수 있다.

[0043] [0058] 트레이닝 종료 조건은, 드리프트 데이터(144) 모두가 적어도 한번 SEC 모델(208)에 제공되었을 때, 특정 수의 트레이닝 반복들이 수행된 후에, 수렴 메트릭이 수렴 임계치를 충족할 때, 또는 트레이닝의 종료를 표시하는 어떤 다른 조건이 충족된 때, 충족될 수 있다. 트레이닝 종료 조건이 충족될 때, 모델 업데이터(152)는 업데이트된 SEC 모델(214)을 저장하며, 여기서 업데이트된 SEC 모델(214)은 트레이닝 동안 적용된 업데이트된 링크 가중치들(212)에 기초한 링크 가중치들을 갖는 SEC 모델(208)에 대응한다.

[0044] [0059] 도 3은 특정 예에 따라 미지의 데이터를 처리하기 위해 SEC 모델(310)을 업데이트하는 양상들을 예시하는 도면이다. 도 3의 SEC 모델(310)은 미지의 데이터(146)와 연관된 도 1의 이용 가능한 SEC 모델들(114) 중 특정 모델을 포함하거나 그에 대응한다. 예컨대, SEC 엔진(120)이 SEC 모델(112)의 출력에 대한 응답으로 트리거 신호(126)를 생성했다면, 미지의 데이터(146)는 SEC 모델(112)과 연관되고, SEC 모델(310)은 SEC 모델(112)에 대응하거나 SEC 모델(112)을 포함한다. 다른 예로서, SEC 엔진(120)이 SEC 모델(118)의 출력에 대한 응답으로 트리거 신호(126)를 생성했다면, 미지의 데이터(146)는 제K SEC 모델(118)과 연관되고, SEC 모델(310)은 제K SEC 모델(118)에 대응하거나 SEC 모델(118)을 포함한다.

[0045] [0060] 도 3의 예에서, 모델 업데이터(152)는 업데이트 모델(306)을 생성한다. 업데이트 모델(306)은 업데이트될 SEC 모델(310), 증분 모델(308) 및 하나 이상의 어댑터 네트워크들(312)을 포함한다. 증분 모델(308)은 SEC 모델(310)과 상이한 출력 계층을 갖는 SEC 모델(310)의 사본이다. 특히, 증분 모델(308)의 출력 계층은 SEC 모델(310)의 출력 계층보다 더 많은 출력 노드들을 포함한다. 예컨대, SEC 모델(310)의 출력 계층은 제1 개수의 노드들(예컨대, N개의 노드들, 여기서 N은 SEC 모델(310)이 트레이닝되어 인식할 사운드 클래스들의 수에 대응

하는 양의 정수임)을 포함하고, 증분 모델(308)의 출력 계층은 제2 개수의 노드들(예컨대, N+M개의 노드들, 여기서 M은, SEC 모델(310)이 인식하도록 트레이닝되지 않은, 업데이트된 SEC 모델(324)이 트레이닝되어 인식할 다수의 새로운 사운드 클래스들에 대응하는 양의 정수임)을 포함한다. 노드들의 제1 개수는 SEC 모델(310)이 트레이닝되어 인식할 사운드 클래스들의 제1 세트의 사운드 클래스들의 개수에 대응하며(예컨대, 사운드 클래스들의 제1 세트는 SEC 모델(310)이 인식할 수 있는 N개의 별개의 사운드 클래스들을 포함함), 노드들의 제2 개수는, 업데이트된 SEC 모델(324)이 트레이닝되어 인식할 사운드 클래스들의 제2 세트의 사운드 클래스들의 개수에 대응한다(예컨대, 사운드 클래스들의 제2 세트는 업데이트된 SEC 모델(324)이 트레이닝되어 인식할 N+M개의 개개의 사운드 클래스들을 포함함). 제2 세트의 사운드 클래스들은 제1 세트의 사운드 클래스들(예컨대, N개의 클래스들) 플러스 하나 이상의 추가적인 사운드 클래스들(예컨대, M개의 클래스들)을 포함한다. 증분 모델(308)의 모델 파라미터들(예컨대, 링크 가중치들)은 SEC 모델(310)의 모델 파라미터들과 동일하도록 초기화된다.

[0046] [0061] 어댑터 네트워크(들)(312)는 뉴럴 어댑터 및 병합 어댑터를 포함한다. 신경 어댑터는 SEC 모델(310)로부터 입력을 수신하도록 그리고 증분 모델(308)의 출력과 병합될 수 있는 출력을 생성하도록 구성된 하나 이상의 어댑터 계층들을 포함한다. 예컨대, SEC 모델(310)은 제1 세트의 사운드 클래스들 중 제1 개수의 클래스들에 대응하는 제1 출력을 생성한다. 특정 양상에서, 제1 출력은 SEC 모델(310)의 출력 계층의 각각의 노드에 대한 하나의 데이터 엘리먼트(예컨대, N개의 데이터 엘리먼트들)를 포함한다. 대조적으로, 증분 모델(308)은 제2 세트의 사운드 클래스들 중 제2 개수의 클래스들에 대응하는 제2 출력을 생성한다. 예컨대, 제2 출력은 증분 모델(308)의 출력 계층의 각각의 노드에 대한 하나의 데이터 엘리먼트(예컨대, N+M개의 데이터 엘리먼트들)를 포함한다. 이 예에서, 어댑터 네트워크(들)(312)의 어댑터 계층(들)은 SEC 모델(310)의 출력을 입력으로서 수신하고, 제2 개수의 데이터 엘리먼트들(예컨대, N+M)을 갖는 출력을 생성한다. 특정 예에서, 어댑터 네트워크(들)(312)의 어댑터 계층(들)은 2개의 완전히 연결된 계층들(예컨대, 입력 계층은 N개의 노드들을 포함하고, 출력 계층은 N+M개의 노드들을 포함하며, 입력 계층의 각각의 노드는 출력 계층의 모든 각각의 노드에 연결됨)을 포함한다.

[0047] [0062] 어댑터 네트워크(들)(312)의 병합 어댑터는 어댑터 계층(들)의 출력과 증분 모델(308)의 출력을 병합함으로써 업데이트 모델(306)의 출력(314)을 생성하도록 구성된다. 예컨대, 병합 어댑터는 어댑터 계층(들)의 출력과 증분 모델(308)의 출력을 엘리먼트별 방식으로 결합하여 결합된 출력을 생성하고 활성화 함수(이를테면, 시그모이드 함수)를 결합된 출력에 에 적용하여 출력(314)을 생성한다. 출력(314)은 업데이트 모델(306)에 의해 오디오 데이터 샘플들(304)이 할당될 사운드 클래스, 업데이트 모델(306)에 의해 결정된 하나 이상의 신뢰도 메트릭 값들, 또는 둘 모두를 표시한다.

[0048] [0063] 모델 업데이터(152)는 오디오 데이터 샘플들(304)에 대응하는 라벨(150) 및 출력(314)을 사용하여, 증분 모델(308), 어댑터 네트워크(들)(312) 또는 둘 모두에 대한 업데이트된 링크 가중치들(316)을 결정한다. SEC 모델(310)의 링크 가중치들은 트레이닝 동안 변하지 않는다. 트레이닝 프로세스는 트레이닝 종료 조건이 충족될 때까지 되풀이하여 반복된다. 트레이닝 동안, 미지의 데이터(146)의 엔트리들 각각은 업데이트 모델(306)에 제공될 수 있다(예컨대, 반복 당 하나의 엔트리). 부가적으로, 일부 구현들에서, 다른 오디오 데이터 샘플들(예컨대, SEC 모델(310)을 트레이닝시키기 위해 이전에 사용된 오디오 데이터 샘플들)은 또한, 증분 모델(308)이 SEC 모델(310)의 이전 트레이닝을 망각할 가능성을 감소시키기 위해 업데이트 모델(306)에 제공될 수 있다.

[0049] [0064] 트레이닝 종료 조건은, 미지의 데이터(146) 모두가 적어도 한번 업데이트 모델(306)에 제공되었을 때, 특정 수의 트레이닝 반복들이 수행된 후에, 수렴 메트릭이 수렴 임계치를 충족할 때, 또는 트레이닝의 종료를 표시하는 어떤 다른 조건이 충족된 때, 충족될 수 있다. 트레이닝 종료 조건이 충족될 때, 모델 체커(320)는 증분 모델(308)과 업데이트 모델(306) 사이에서 업데이트된 SEC 모델(324)(예컨대, SEC 모델(310), 증분 모델(308) 및 어댑터 네트워크(들)(312)의 결합)을 선택한다.

[0050] [0065] 특정 양상에서, 모델 체커(320)는 증분 모델(308)에 의해 할당된 사운드 클래스들(322)의 정확도 및 SEC 모델(310)에 의해 할당된 사운드 클래스들(322)의 정확도에 기반하여 업데이트된 SEC 모델(324)을 선택한다. 예컨대, 모델 체커(320)는 (증분 모델(308)에 의해 할당된 사운드 클래스들(322)에 기반하여) 증분 모델(308)에 대한 F1-스코어 및 (SEC 모델(310)에 의해 할당된 사운드 클래스들(322)에 기반하여) SEC 모델(310)의 F1-스코어를 결정할 수 있다. 이 예에서, 증분 모델(308)의 F1-스코어의 값이 SEC 모델(310)의 F1-스코어의 값보다 크거나 같으면, 모델 체커(320)는 업데이트된 SEC 모델(324)로서 증분 모델(308)을 선택한다. 일부 구현들에서, 모델 체커(320)는, 증분 모델(308)의 F1-스코어의 값이 SEC 모델(310)의 F1-스코어의 값 이상이면(또는 임계량 미만만큼 SEC 모델(310)의 F1-스코어의 값 미만이면), 업데이트된 SEC 모델(324)로서 증분 모델(308)을 선택한

다. 증분 모델(308)의 F1-스코어의 값이 SEC 모델(310)에 대한 F1-스코어의 값보다 작으면(또는 SEC 모델(310)에 대한 F1-스코어의 값보다 임계치 초과만큼 더 작으면), 모델 체커(320)는 업데이트된 SEC 모델(324)로서 업데이트 모델(306)을 선택한다. 업데이트된 SEC 모델(324)로서 증분 모델(308)이 선택되면, SEC 모델(310), 어댑터 네트워크(들)(312) 또는 둘 모두가 폐기될 수 있다.

[0051] [0066] 일부 구현들에서, 모델 체커(320)는 모델 업데이터(152)와 통합되거나 생략된다. 예컨대, 업데이트 모델(306)을 트레이닝시킨 후, 업데이트 모델(306)은 (예컨대, 업데이트 모델(306)과 증분 모델(308) 사이에 어떠한 선택도 없이) 업데이트된 SEC 모델(324)로서 저장될 수 있다. 예로서, 업데이트 모델(306)을 트레이닝하는 동안, 모델 업데이터(152)는 증분 모델(308)에 대한 정확도 메트릭을 결정할 수 있다. 이러한 예에서, 트레이닝 종료 조건은, 트레이닝 후에, (예컨대, 업데이트 모델(306)과 증분 모델(308) 사이에 어떠한 선택도 없이) 증분 모델(308)이 업데이트된 SEC 모델(324)로서 저장되도록 증분 모델(308)에 대한 정확도 메트릭에 기반할 수 있다.

[0052] [0067] 도 3을 참조하여 설명된 전이 러닝 기법들을 활용하여, 모델 체커(320)는 도 1의 디바이스(100)가 이전까지 미지의 사운드 클래스를 인식하도록 SEC 모델을 업데이트할 수 있게 한다. 추가적으로, 설명된 전이 러닝 기법들은 이전까지 미지의 사운드 클래스를 인식하기 위해 SEC 모델을 처음부터 트레이닝시키는 데 사용될 것보다 상당히 더 적은 컴퓨터 자원들(예컨대, 메모리, 프로세싱 시간 및 전력)을 사용한다.

[0053] [0068] 일부 구현들에서, 도 2를 참조하여 설명된 동작들(예컨대, 드리프트 데이터(144)에 기반하여 업데이트된 SEC 모델(214)을 생성하는 것)은 도 1의 디바이스(100)에서 수행되고, 도 3을 참조하여 설명된 동작들(예컨대, 미지의 데이터(146)에 기반하여 업데이트된 SEC 모델(324)을 생성하는 것)은 상이한 디바이스(이를테면, 도 6의 원격 컴퓨팅 디바이스(618))에서 수행된다. 예시하자면, 미지의 데이터(146) 및 라벨(들)(150)은 디바이스(100)에서 캡처되어 더 많은 이용 가능한 컴퓨팅 자원들을 갖는 제2 디바이스에 송신될 수 있다. 이 예에서, 제2 디바이스는 업데이트된 SEC 모델(324)을 생성하고, 디바이스(100)는 제2 디바이스로부터 업데이트된 SEC 모델(324)을 나타내는 송신 또는 데이터를 다운로드하거나 수신한다. 미지의 데이터(146)에 기반하여 업데이트된 SEC 모델(324)을 생성하는 것은 드리프트 데이터(144)에 기반하여 업데이트된 SEC 모델(214)을 생성하는 것보다 더 자원 집약적인 프로세스이다(예컨대, 더 많은 메모리, 전력 및 프로세서 시간을 사용함). 따라서, 도 2를 참조하여 설명된 동작들 및 도 3을 참조하여 설명된 동작들을 상이한 디바이스들 사이에서 분할하는 것은 디바이스(100)의 자원들을 보존할 수 있다.

[0054] [0069] 도 4는 도 1의 디바이스의 동작의 특정 예를 예시하는 도면이다. 도 4는, 활성 SEC 모델(예컨대, SEC 모델(112))이, 오디오 데이터 샘플들(110)이 캡처되는 오디오 장면에 대응하는지 여부에 대한 결정이, 현재 오디오 장면을 이전 오디오 장면과 비교하는 것에 기반하는 구현을 예시한다.

[0055] [0070] 도 4에서, 마이크로폰(들)(104)에 의해 캡처된 오디오 데이터는 오디오 데이터 샘플들(110)을 생성하는 데 사용된다. 오디오 데이터 샘플들(110)은 오디오 분류(402)를 수행하는 데 사용된다. 예컨대, 이용 가능한 SEC 모델들(114) 중 하나 이상이 도 1의 SEC 엔진(120)에 의해 활성 SEC 모델로서 사용된다. 특정 양상에서, 활성 SEC 모델은, 이전 오디오 장면(408)으로 또한 지칭되는 이전 샘플링 기간 동안 장면 검출기(140)에 의해 표시된 오디오 장면에 기반하여, 이용 가능한 SEC 모델들(114) 중에서 선택된다.

[0056] [0071] 오디오 분류(402)는 활성 SEC 모델을 사용하여 오디오 데이터 샘플들(110)의 분석에 기반하여 결과(404)를 생성한다. 결과(404)는 오디오 데이터 샘플들(110)과 연관된 사운드 클래스, 오디오 데이터 샘플들(110)이 특정 사운드 클래스에 대응하는 확률, 또는 오디오 데이터 샘플들(110)의 사운드 클래스가 미지일 확률을 표시할 수 있다. 결과(404)가 오디오 데이터 샘플들(110)이 알려진 사운드 클래스에 대응한다고 표시하면, 블록(406)에서, 오디오 데이터 샘플들(110)과 연관된 사운드 클래스(122)를 나타내는 출력(124)을 생성하라는 결정이 행해진다. 예컨대, 도 1의 SEC 엔진(120)은 출력(124)을 생성할 수 있다.

[0057] [0072] 결과(404)가 오디오 데이터 샘플들(110)이 알려진 사운드 클래스에 대응하지 않는다고 표시하면, 블록(406)에서, 트리거(126)를 생성하라는 결정이 행해진다. 트리거(126)는 드리프트 검출 방식을 활성화시키며, 도 4에서, 이는 장면 검출기(140)로 하여금 센서(들)(134)로부터의 데이터에 기반하여 현재 오디오 장면(407)을 식별하게 하는 것을 포함한다.

[0058] [0073] 활성 SEC 모델이 선택된 이후 오디오 장면 변화가 발생하였는지 여부를 결정하기 위해, 현재 오디오 장면(407)이 블록(410)에서 이전 오디오 장면(408)과 비교된다. 블록(412)에서, 오디오 데이터 샘플들(110)의 사운드 클래스가 드리프트로 인해 인식되지 않았는지 여부의 결정이 행해진다. 예컨대, 현재 오디오 장면(407)이

이전 오디오 장면(408)에 대응하지 않으면, 블록(412)에서의 결정은, 오디오 데이터 샘플들(110)의 사운드 클래스가 인식되지 않는 원인이 드리프트가 아니었다는 것이다. 이러한 상황에서, 오디오 데이터 샘플들(110)은 폐기되거나, 또는 블록(414)에서 미지의 데이터로서 저장될 수 있다.

[0059] [0074] 현재 오디오 장면(407)이 이전 오디오 장면(408)에 대응하면, 블록(412)에서의 결정은, 활성 SEC 모델이 현재 오디오 장면(407)에 대응하기 때문에 오디오 데이터 샘플들(110)의 사운드 클래스가 드리프트로 인해 인식되지 않았다는 것이다. 이러한 상황에서, 드리프트된 사운드 클래스가 블록(416)에서 식별되고, 그리고 오디오 데이터 샘플들(110) 및 사운드 클래스의 식별자는 블록(418)에서 드리프트 데이터로서 저장된다.

[0060] [0075] 충분한 드리프트 데이터가 저장될 때, SEC 모델은 블록(420)에서 업데이트되어, 업데이트된 SEC 모델(154)을 생성한다. 업데이트된 SEC 모델(154)은 이용 가능한 SEC 모델들(114)에 추가된다. 일부 구현들에서, 업데이트된 SEC 모델(154)은 결과(404)를 생성한 활성 SEC 모델을 대체한다.

[0061] [0076] 도 5는 도 1의 디바이스의 동작의 다른 특정 예를 예시하는 도면이다. 도 5는, 활성 SEC 모델(예컨대, SEC 모델(112))이, 오디오 데이터 샘플들(110)이 캡처되는 오디오 장면에 대응하는지 여부에 대한 결정이, 현재 오디오 장면을 활성 SEC 모델을 설명하는 정보와 비교하는 것에 기반하는 구현을 예시한다.

[0062] [0077] 도 5에서, 마이크로폰(들)(104)에 의해 캡처된 오디오 데이터는 오디오 데이터 샘플들(110)을 생성하는 데 사용된다. 오디오 데이터 샘플들(110)은 오디오 분류(402)를 수행하는 데 사용된다. 예컨대, 이용 가능한 SEC 모델들(114) 중 하나 이상이 도 1의 SEC 엔진(120)에 의해 활성 SEC 모델로서 사용된다. 특정 양상에서, 활성 SEC 모델은 이용 가능한 SEC 모델들(114) 중에서 선택된다. 일부 구현들에서, 이용 가능한 SEC 모델들(114) 중 하나 이상을 활성 SEC 모델로서 선택하기보다는, 이용 가능한 SEC 모델들(114)의 앙상블이 사용된다.

[0063] [0078] 오디오 분류(402)는 하나 이상의 이용 가능한 SEC 모델(114)을 사용하여 오디오 데이터 샘플들(110)의 분석에 기반하여 결과(404)를 생성한다. 결과(404)는 오디오 데이터 샘플들(110)과 연관된 사운드 클래스, 오디오 데이터 샘플들(110)이 특정 사운드 클래스에 대응하는 확률, 또는 오디오 데이터 샘플들(110)의 사운드 클래스가 미지일 확률을 표시할 수 있다. 결과(404)가 오디오 데이터 샘플들(110)이 알려진 사운드 클래스에 대응한다고 표시하면, 블록(406)에서, 오디오 데이터 샘플들(110)과 연관된 사운드 클래스(122)를 나타내는 출력(124)을 생성하라는 결정이 행해진다. 예컨대, 도 1의 SEC 엔진(120)은 출력(124)을 생성할 수 있다.

[0064] [0079] 결과(404)가 오디오 데이터 샘플들(110)이 알려진 사운드 클래스에 대응하지 않는다고 표시하면, 블록(406)에서, 트리거(126)를 생성하라는 결정이 행해진다. 트리거(126)는 드리프트 검출 방식을 활성화시키며, 이는 도 5에서, 장면 검출기(140)로 하여금 센서(들)(134)로부터의 데이터에 기반하여 현재 오디오 장면을 식별하게 하고, 트리거(126)를 전송하게 한 결과(404)를 생성한 SEC 모델에 현재 오디오 장면이 대응하는지 여부를 결정하게 하는 것을 포함한다.

[0065] [0080] 블록(412)에서, 오디오 데이터 샘플들(110)의 사운드 클래스가 드리프트로 인해 인식되지 않았는지 여부의 결정이 행해진다. 예컨대, 현재 오디오 장면이 결과(404)를 생성한 SEC 모델에 대응하지 않으면, 블록(412)에서의 결정은, 오디오 데이터 샘플들(110)의 사운드 클래스가 인식되지 않는 원인이 드리프트가 아니었다는 것이다. 이러한 상황에서, 오디오 데이터 샘플들(110)은 폐기되거나, 또는 블록(414)에서 미지의 데이터로서 저장될 수 있다.

[0066] [0081] 현재 오디오 장면이 결과(404)를 생성한 SEC 모델에 대응하면, 블록(412)에서의 결정은, 오디오 데이터 샘플들(110)의 사운드 클래스가 드리프트로 인해 인식되지 않았다는 것이다. 이러한 상황에서, 드리프트된 사운드 클래스가 블록(416)에서 식별되고, 그리고 오디오 데이터 샘플들(110) 및 사운드 클래스의 식별자는 블록(418)에서 드리프트 데이터로서 저장된다.

[0067] [0082] 충분한 드리프트 데이터가 저장될 때, SEC 모델은 블록(420)에서 업데이트되어, 업데이트된 SEC 모델(154)을 생성한다. 업데이트된 SEC 모델(154)은 이용 가능한 SEC 모델들(114)에 추가된다. 일부 구현들에서, 업데이트된 SEC 모델(154)은 결과(404)를 생성한 활성 SEC 모델을 대체한다.

[0068] [0083] 도 6은 도 1의 디바이스(100)의 특정 예를 예시하는 블록도이다. 도 6에서, 디바이스(100)는 오디오 데이터 샘플들(예컨대, 도 1의 오디오 데이터 샘플들(110))의 입력에 대한 응답으로 사운드 식별 데이터(예컨대, 도 1의 출력(124))를 생성하기 위해 SEC 모델들(114)을 사용하도록 구성된다. 부가적으로, 도 6의 디바이스(100)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나 이상을 업데이트하도록 구성된다. 예컨대, 디바이스(100)는 도 2를 참조하여 설명된 바와 같이 드리프트 데이터(144)를 사용하여 SEC 모델들(114)을 업데이트하도록 구성되거나, 도 3을 참조하여 설명된 바와 같이 미지의 데이터(146)를 사용하여 SEC 모델들

(114)을 업데이트하도록 구성되거나, 또는 둘 모두일 수 있다. 일부 구현들에서, 원격 컴퓨팅 디바이스(618)는 일부 상황들에서 SEC 모델들(114)을 업데이트한다. 예시하자면, 디바이스(100)는 드리프트 데이터(144)를 사용하여 SEC 모델들(114)을 업데이트할 수 있고, 원격 컴퓨팅 디바이스(618)는 미지의 데이터(146)를 사용하여 SEC 모델들(114)을 업데이트할 수 있다. 다양한 구현들에서, 디바이스(100)는 도 6에 예시된 것보다 많거나 적은 컴포넌트들을 가질 수 있다.

[0069] [0084] 특정 구현에서, 디바이스(100)는 프로세서(604)(예컨대, CPU(central processing unit))를 포함한다. 디바이스(100)는 하나 이상의 추가적인 프로세서(들)(606)(예컨대, 하나 이상의 디지털 신호 프로세서(DSP)들)를 포함할 수 있다. 프로세서(604), 프로세서(들)(606) 또는 둘 모두는 사운드 식별 데이터를 생성하도록, SEC 모델(114)을 업데이트하도록, 또는 둘 모두를 수행하도록 구성될 수 있다. 예컨대, 도 6에서, 프로세서(들)(606)는 SEC 엔진(120)을 포함한다. SEC 엔진(120)은 SEC 모델들(114) 중 하나 이상을 사용하여 오디오 데이터 샘플들을 분석하도록 구성된다.

[0070] [0085] 도 6에서, 디바이스(100)는 또한 메모리(608) 및 코덱(624)을 포함한다. 메모리(608)는 도 1 - 도 5를 참조하여 설명한 하나 이상의 동작들을 구현하기 위해 프로세서(604) 또는 프로세서(들)(606)에 의해 실행 가능한 명령들(660)을 저장한다. 일 예에서, 명령들(660)은 특징 추출기(108), SEC 엔진(120), 장면 검출기(140), 드리프트 검출기(128), 모델 업데이터(152), 모델 체커(320), 또는 이들의 조합을 포함하거나 그에 대응한다. 메모리(608)는 또한 설정 데이터(130), SEC 모델들(114), 및 모델 업데이트 데이터(148)를 저장할 수 있다.

[0071] [0086] 도 6에서, 스피커(들)(622) 및 마이크로폰(들)(104)은 CODEC(624)에 결합될 수 있다. 특정 양상에서, 마이크로폰(들)(104)은 디바이스(100)와 연관된 음향 환경을 표현하는 오디오를 수신하도록, 그리고 특징 추출기가 오디오 데이터 샘플들을 생성하기 위해 사용하는 오디오 신호들을 생성하도록 구성된다. 도 6에 예시된 예에서, 코덱(CODEC)(624)은 DAC(digital-to-analog converter)(626)) 및 ADC(analog-to-digital converter)(628))를 포함한다. 특정 구현에서, 코덱(624)은 마이크로폰(들)(104)으로부터 아날로그 신호들을 수신하고, ADC(628)를 사용하여 아날로그 신호들을 디지털 신호들로 변환하고, 디지털 신호들을 프로세서(들)(606)에 제공한다. 특정 구현에서, 프로세서(들)(606)는 디지털 신호들을 코덱(624)에 제공하고, 코덱(624)은 DAC(626)를 사용하여 디지털 신호들을 아날로그 신호들로 변환하고 아날로그 신호들을 스피커(들)(622)에 제공한다.

[0072] [0087] 도 6에서, 디바이스(100)는 또한 입력 디바이스(136)를 포함한다. 디바이스(100)는 또한 디스플레이 제어기(610)에 결합된 디스플레이(620)를 포함할 수 있다. 특정 양상에서, 입력 디바이스(136)는 센서, 키보드, 포인팅 디바이스 등을 포함한다. 일부 구현들에서, 입력 디바이스(136) 및 디스플레이(620)는 터치스크린 또는 유사한 터치 또는 모션 감지 디스플레이로 결합된다. 입력 디바이스(136)는 트레이닝 데이터(302)를 생성하기 위해 미지의 데이터(146)와 연관된 라벨을 제공하는 데 사용될 수 있다. 입력 디바이스(136)는 또한, 이를테면, 도 2를 참조하여 설명된 모델 업데이트 프로세스 또는 도 3을 참조하여 설명된 모델 업데이트 프로세스를 시작하기 위해 모델 업데이트 동작을 개시하는 데 사용될 수 있다. 일부 구현들에서, 입력 디바이스(136)는 또한, 또는 대안적으로, SEC 엔진(120)에 의해 사용될, 이용 가능한 SEC 모델들(114)의 특정 SEC 모델을 선택하는 데 사용될 수 있다. 특정 양상에서, 입력 디바이스(136)는 SEC 엔진(120)에 의해 사용될 SEC 모델을 선택하는 데 사용될 수 있는 설정 데이터(130)를 구성하기 위해, 오디오 장면(142)을 결정하기 위해, 또는 둘 모두를 위해 사용될 수 있다. 디스플레이(620)는, SEC 모델들 중 하나(예컨대, 도 1의 출력(124))에 의한 분석의 결과들을 디스플레이하기 위해, 미지의 데이터(146)와 연관된 라벨을 제공하도록 사용자에게 프롬프트를 디스플레이 하기 위해, 또는 둘 모두를 위해 사용될 수 있다.

[0073] [0088] 일부 구현들에서, 디바이스(100)는 또한 트랜시버(614)에 결합된 모뎀(612)을 포함한다. 도 6에서, 트랜시버(614)는 원격 컴퓨팅 디바이스(618)와 같은 다른 디바이스들과의 무선 통신을 가능하게 하기 위해 안테나(616)에 결합된다. 다른 예들에서, 트랜시버(614)는 또한 또는 대안적으로, 원격 컴퓨팅 디바이스(618)와 같은 다른 디바이스들과의 유선 통신을 가능하게 하기 위해 통신 포트(예컨대, 이더넷 포트)에 결합된다.

[0074] [0089] 도 6에서, 디바이스(100)는 클록(132) 및 센서들(134)을 포함한다. 특정 예들로서, 센서들(134)은 하나 이상의 카메라들(650), 하나 이상의 포지션 센서들(652), 마이크로폰(들)(104), 다른 센서(들)(654), 또는 이들의 조합을 포함한다.

[0075] [0090] 특정 양상에서, 클록(132)은 특정 오디오 데이터 샘플들이 수신된 때를 표시하기 위해 특정 오디오 데이터 샘플들에 타임스탬프를 할당하는 데 사용될 수 있는 클록 신호를 생성한다. 이러한 양상에서, SEC 엔진(120)은 특정 오디오 데이터 샘플들을 분석하는 데 사용할 특정 SEC 모델(114)을 선택하기 위해 타임스탬프를 사용

할 수 있다. 추가적으로 또는 대안적으로, 타임스탬프는 특정 오디오 데이터 샘플들과 연관된 오디오 장면(142)을 결정하기 위해 장면 검출기(140)에 의해 사용될 수 있다.

[0076] [0091] 특정 양상에서, 카메라(들)(650)는 이미지의 데이터, 비디오 데이터 또는 둘 모두를 생성한다. SEC 엔진(120)은 오디오 데이터 샘플들을 분석하는 데 사용할 특정 SEC 모델(114)을 선택하기 위해 이미지의 데이터, 비디오 데이터 또는 둘 모두를 사용할 수 있다. 추가적으로 또는 대안적으로, 이미지 데이터, 비디오 데이터 또는 둘 모두는 특정 오디오 데이터 샘플들과 연관된 오디오 장면(142)을 결정하기 위해 장면 검출기(140)에 의해 사용될 수 있다. 예컨대, 특정 SEC 모델(114)은 실외 사용을 위해 지정될 수 있고, 디바이스(100)가 실외 환경에 로케이팅되었음을 확인하기 위해 이미지 데이터, 비디오 데이터 또는 둘 모두가 사용될 수 있다.

[0077] [0092] 특정 양상에서, 포지션 센서(들)(652)는 디바이스(100)의 로케이션을 나타내는 글로벌 포지션 데이터와 같은 포지션 데이터를 생성한다. SEC 엔진(120)은 오디오 데이터 샘플들을 분석하는 데 사용할 특정 SEC 모델(114)을 선택하기 위해 포지션 데이터를 사용할 수 있다. 추가적으로 또는 대안적으로, 포지션 데이터는 특정 오디오 데이터 샘플들과 연관된 오디오 장면(142)을 결정하기 위해 장면 검출기(140)에 의해 사용될 수 있다. 예컨대, 특정 SEC 모델(114)은 집에서 사용하도록 지정될 수 있고, 포지션 데이터는 디바이스(100)가 집 로케이팅에 로케이팅되었음을 확인하는 데 사용될 수 있다. 포지션 센서(들)(652)는 위성-기반 포지셔닝 시스템을 위한 수신기, 로컬 포지셔닝 시스템 수신기를 위한 수신기, 관성 내비게이션 시스템, 랜드마크-기반 포지셔닝 시스템, 또는 이들의 조합을 포함할 수 있다.

[0078] [0093] 다른 센서(들)(654)는 예컨대, 배향 센서, 자력계, 광 센서, 접촉 센서, 온도 센서 또는 디바이스(100)에 결합되거나 디바이스(100) 내에 포함될 수 있고, 특정 시간에 디바이스(100)와 연관된 오디오 장면(142)을 결정하는 데 유용한 장면 데이터(138)를 생성하는 데 사용될 수 있는 임의의 다른 센서를 포함할 수 있다.

[0079] [0094] 특정 구현에서, 디바이스(100)는 시스템-인-패키지 또는 시스템-온-칩 디바이스(602)에 포함된다. 특정 구현에서, 메모리(608), 프로세서(604), 프로세서(들)(606), 디스플레이 제어기(610), 코덱(624), 모뎀(612) 및 트랜시버(614)는 시스템-인-패키지 또는 시스템-온-칩 디바이스(602)에 포함된다. 특정한 구현에서, 입력 디바이스(136) 및 전력 공급부(630)는 시스템-온-칩 디바이스(602)에 커플링된다. 또한, 특정한 구현에서, 도 6에 예시된 바와 같이, 디스플레이(620), 입력 디바이스(136), 스피커(들)(622), 센서들(134), 클록(132), 안테나(616), 및 전력 공급부(630)는 시스템-온-칩 디바이스(602) 외부에 있다. 특정 구현에서, 디스플레이(620), 입력 디바이스(136), 스피커(들)(622), 센서들(134), 클록(132), 안테나(616) 및 전력 공급부(630) 각각은 인터페이스 또는 제어기와 같은 시스템-온-칩 디바이스(602)의 컴포넌트에 결합될 수 있다.

[0080] [0095] 디바이스(100)는 음성 활성화 디바이스, 오디오 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 차량, 컴퓨팅 디바이스, 통신 디바이스, 사물 인터넷(IoT) 디바이스, 가상 현실(VR) 디바이스, 증강 현실(AR) 디바이스, 혼합 현실(MR) 디바이스, 스마트 스피커, 모바일 컴퓨팅 디바이스, 모바일 통신 디바이스, 스마트 전화, 휴대 전화, 랩톱 컴퓨터, 컴퓨터, 태블릿, 개인용 디지털 보조기, 디스플레이 디바이스, 텔레비전, 게임 콘솔, 기기, 음악 플레이어, 라디오, 디지털 비디오 플레이어, 디지털 비디오 디스크(DVD) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 또는 이들의 임의의 조합을 포함하거나, 이에 대응하거나, 이에 포함될 수 있다. 특정 양상에서, 프로세서(604), 프로세서(들)(606) 또는 이들의 조합은 집적 회로에 포함된다.

[0081] [0096] 도 7은 도 1의 디바이스(100)의 양상들을 통합하는 차량(700)의 예시적인 예이다. 일 구현에 따르면, 차량(700)은 자율 주행 자동차이다. 다른 구현들에 따르면, 차량(700)은 자동차, 트럭, 모터사이클, 항공기, 수상 차량 등이다. 도 7에서, 운송수단(700)은 디스플레이(620), 센서들(134) 중 하나 이상, 디바이스(100), 또는 이들의 조합을 포함한다. 센서들(134) 및 디바이스(100)는, 이들 컴포넌트들이 차량(700)의 승객들에게 보이지 않을 수 있음을 표시하기 위해 점선을 사용하여 도시된다. 디바이스(100)는 차량(700)에 통합되거나 차량(700)에 결합될 수 있다.

[0082] [0097] 특정 양상에서, 디바이스(100)는 디스플레이(620)에 결합되고, 본원에서 설명된 다양한 이벤트들(예컨대, 사운드 이벤트들)을 검출 또는 인식하는 SEC 모델들(114) 중 하나에 대한 응답으로 디스플레이(620)에 출력을 제공한다. 예컨대, 디바이스(100)는 마이크로폰(들)(104)으로부터 수신된 오디오 데이터 샘플들(110)에서 SEC 모델들(114) 중 하나에 의해 인식된 사운드(102)(이를테면, 자동차 경적음)의 사운드 클래스를 표시하는, 도 1의 출력(124)을 디스플레이(620)에 제공한다. 일부 구현들에서, 디바이스(100)는 사운드 이벤트를 인식하는 것에 대한 응답으로, 차량의 조작자에게 경고하는 것 또는 센서들(134) 중 하나를 활성화시키는 것과 같은 액션을 수행할 수 있다. 특정 예에서, 디바이스(100)는 인식된 사운드 이벤트에 대한 응답으로 액션이

수행되고 있는지 여부를 표시하는 출력을 제공한다. 특정 양상에서, 사용자는 인식된 사운드 이벤트들에 대한 응답으로 액션들의 수행을 인에이블 또는 디스에이블하기 위해 디스플레이(620) 상에 디스플레이되는 옵션을 선택할 수 있다.

[0083] [0098] 특정 구현들에서, 센서들(134)은 도 1의 마이크로폰(들)(104), 차량 점유 센서들, 눈 추적 센서, 포지션 센서(들)(652), 또는 외부 환경 센서들(예컨대, 라이다 센서들 또는 카메라들)을 포함한다. 특정 양상에서, 센서들(134)의 센서 입력은 사용자의 로케이션을 표시한다. 예컨대, 센서들(134)은 차량(700) 내의 다양한 로케이션들과 연관된다.

[0084] [0099] 도 7의 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 차량(700)에 설치되거나 차량(700)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 차량(700)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 차량(700)에 설치되거나 또는 차량(700)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0085] [0100] 따라서, 도 1 - 도 5와 관련하여 설명된 기법들은 차량(700)의 사용자가, 디바이스(100)가 동작하는 음향 환경에 특정한(어쩌면 고유한) 드리프트를 처리하기 위해 업데이트된 SEC 모델을 생성할 수 있게 한다. 일부 구현들에서, 디바이스(100)는 사용자 개입 없이 업데이트된 SEC 모델을 생성할 수 있다. 또한, 도 1 - 도 5와 관련하여 설명된 기법들은 차량(700)의 사용자가, 하나 이상의 새로운 사운드 클래스들을 검출하기 위해 업데이트된 SEC 모델을 생성할 수 있게 한다. 또한, SEC 모델은 차량(700)에 탑재된 컴퓨팅 자원들의 과도한 사용 없이 업데이트될 수 있다. 예컨대, 차량(700)은 SEC 모델을 처음부터 트레이닝하는 데 사용된 모든 트레이닝 데이터를 로컬 메모리에 저장할 필요가 없다.

[0086] [0101] 도 8은 헤드셋(802), 이를테면 가상 현실 헤드셋, 증강 현실 헤드셋, 혼합 현실 헤드셋, 확장 현실 헤드셋, 머리-장착 디스플레이, 또는 이들의 조합에 결합되거나 헤드셋(802) 내에 통합된 디바이스(100)의 예를 도시한다. 시각적 인터페이스 디바이스, 이를테면 디스플레이(620)는, 헤드셋(802)이 착용된 동안 사용자에게 증강 현실, 혼합 현실 또는 가상 현실 이미지들 또는 장면들의 디스플레이를 가능하게 하기 위해 사용자의 눈들 앞에 포지셔닝된다. 특정 예에서, 디스플레이(620)는 인식된 사운드 이벤트의 표시와 같은 디바이스(100)의 출력(예컨대, 도 1의 출력(124))을 디스플레이하도록 구성된다. 헤드셋(802)은 센서들(134), 이를테면, 도 1의 마이크로폰(들)(104), 도 6의 카메라(들)(650), 도 6의 포지션 센서(들)(652), 도 6의 다른 센서들(654), 또는 이들의 조합을 포함한다. 단일 로케이션으로 예시되어 있지만, 다른 구현들에서, 센서들(134)은 헤드셋(802)의 다른 로케이션들, 이를테면, 멀티-모달 입력들을 검출하기 위해 헤드셋(802) 주위에 분포된 하나 이상의 마이크로폰들 및 하나 이상의 카메라들의 어레이에 포지셔닝될 수 있다.

[0087] [0102] 센서들(134)은, 디바이스(100)가 사운드 이벤트들을 검출하거나 SEC 모델들(114)을 업데이트하기 위해 사용하는 오디오 데이터의 검출을 가능하게 한다. 예컨대, SEC 엔진(120)은 SEC 모델들(114) 중 하나 이상을 사용하여, 인식된 사운드 이벤트, 이를테면 자동차 경적음이 센서들(134)로부터 수신된 오디오 데이터 샘플들에서 검출되는 것을 표시하기 위해 디스플레이(620)에 제공될 수 있는 사운드 이벤트 분류 데이터를 생성한다. 일부 구현들에서, 디바이스(100)는 사운드 이벤트를 인식하는 것에 대한 응답으로, 카메라 또는 센서들(134) 중 다른 하나를 활성화시키는 것 또는 사용자에게 햅틱 피드백을 제공하는 것과 같은 액션을 수행할 수 있다.

[0088] [0103] 도 8에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 헤드셋(802)에 설치되거나 헤드셋(802)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 헤드셋(802)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 헤드셋(802)에 설치되거나 또는 헤드셋(802)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0089] [0104] 도 9는 디스플레이(620) 및 센서들(134)을 포함하는 "스마트 워치"로서 예시된 웨어러블 전자 디바이스(902)에 통합된 디바이스(100)의 예를 도시한다. 센서들(134)은 예컨대, 위치, 비디오, 스피치 및 제스처와 같은 양상들에 기초한 사용자 입력 및 오디오 장면들의 검출을 가능하게 한다. 센서들(134)은 또한 웨어러블 전자 디바이스(902) 주위의 음향 환경에서 사운드들의 검출을 가능하게 하며, 디바이스(100)는 사운드 이벤트들을 검

출하거나 SEC 모델들(114)을 업데이트하기 위해 이를 사용한다. 예컨대, 디바이스(100)는 인식된 사운드 이벤트가 센서들(134)로부터 수신된 오디오 데이터 샘플들에서 검출됨을 나타내는 도 1의 출력(124)을 디스플레이(620)에 제공한다. 일부 구현들에서, 디바이스(100)는 사운드 이벤트를 인식하는 것에 대한 응답으로, 카메라 또는 센서들(134) 중 다른 하나를 활성화시키는 것 또는 사용자에게 햅틱 피드백을 제공하는 것과 같은 액션을 수행할 수 있다.

[0090] [0105] 도 9에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 웨어러블 전자 디바이스(902)에 설치되거나 웨어러블 전자 디바이스(902)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 웨어러블 전자 디바이스(902)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 웨어러블 전자 디바이스(902)에 설치되거나 또는 웨어러블 전자 디바이스(902)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0091] [0106] 도 10은 음성-제어 스피커 시스템(1000)의 예시적인 예이다. 음성-제어 스피커 시스템(1000)은 무선 네트워크 연결될 수 있고 보조 동작을 실행하도록 구성된다. 도 10에서, 디바이스(100)는 음성-제어 스피커 시스템(1000)에 포함된다. 음성-제어 스피커 시스템(1000)은 또한 스피커(1002) 및 센서들(134)을 포함한다. 센서들(134)은 음성 입력 또는 다른 오디오 입력을 수신하기 위해 도 1의 마이크로폰(들)(104)을 포함한다.

[0092] [0107] 동작 동안, 구두 커맨드 또는 인식된 사운드 이벤트를 수신하는 것에 대한 응답으로, 음성-제어 스피커 시스템(1000)은 보조 동작들을 실행할 수 있다. 보조 동작들은 온도를 조정하는 것, 음악을 재생하는 것, 조명들을 켜는 것 등을 포함할 수 있다. 센서들(134)은, 디바이스(100)가 사운드 이벤트들을 검출하거나 SEC 모델들(114) 중 하나 이상을 업데이트하기 위해 사용하는 오디오 데이터 샘플들의 검출을 가능하게 한다. 추가적으로, 음성-제어 스피커 시스템(1000)은 디바이스(100)에 의해 인식된 사운드 이벤트들에 기반하여 일부 동작들을 실행할 수 있다. 예컨대, 디바이스(100)가 도어 폐쇄의 사운드를 인식하면, 음성-제어 스피커 시스템(1000)은 하나 이상의 조명들을 켤 수 있다.

[0093] [0108] 도 10에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 음성-제어 스피커 시스템(1000)에 설치되거나 음성-제어 스피커 시스템(1000)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 음성-제어 스피커 시스템(1000)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 음성-제어 스피커 시스템(1000)에 설치되거나 또는 음성-제어 스피커 시스템(1000)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0094] [0109] 도 11은 도 1의 디바이스(100)의 양상들을 통합하는 카메라(1100)를 예시한다. 도 11에서, 디바이스(100)는 카메라(1100)에 통합되거나 그에 결합된다. 카메라(1100)는 이미지 센서(1102) 및 하나 이상의 다른 센서들(예컨대, 센서들(134)), 이를테면, 도 1의 마이크로폰(들)(104)을 포함한다. 추가로, 카메라(1100)는 오디오 데이터 샘플들에 기반하여 사운드 이벤트들을 식별하고 SEC 모델들(114) 중 하나 이상을 업데이트하도록 구성된 디바이스(100)를 포함한다. 특정 양상에서, 카메라(1100)는 인식된 사운드 이벤트에 대한 응답으로 하나 이상의 액션들을 수행하도록 구성된다. 예컨대, 카메라(1100)는 디바이스(100)가 센서들(134)로부터의 오디오 데이터 샘플들에서 특정 사운드 이벤트를 검출하는 것에 대한 응답으로 이미지 센서(1102)로 하여금 이미지를 캡처하게 할 수 있다.

[0095] [0110] 도 11에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 카메라(1100)에 설치되거나 헤드셋(1100)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 카메라(1100)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 카메라(1100)에 설치되거나 또는 카메라(1100)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0096] [0111] 도 12는 도 1의 디바이스(100)의 양상들을 통합하는 모바일 디바이스(1200)를 예시한다. 도 12에서, 모

바일 디바이스(1200)는 도 1의 디바이스(100)를 포함하거나 그에 결합된다. 모바일 디바이스(1200)는 예시적인 비제한적인 예들로서 폰 또는 태블릿을 포함한다. 모바일 디바이스(1200)는 디스플레이(620) 및 센서들(134), 이를테면, 도 1의 마이크로폰(들)(104), 도 6의 카메라(들)(650), 도 6의 포지션 센서(들)(652), 또는 도 6의 다른 센서(들)(654)를 포함한다. 동작 동안, 모바일 디바이스(1200)는 디바이스(100)가 특정 사운드 이벤트를 인식하는 것에 대한 응답으로 특정 액션들을 수행할 수 있다. 예컨대, 액션들은 서모스탯, 홈 오토메이션 시스템, 다른 모바일 디바이스 등과 같은 다른 디바이스들에 커맨드들을 전송하는 것을 포함할 수 있다.

[0097] [0112] 도 12에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 모바일 디바이스(1200)에 설치되거나 모바일 디바이스(1200)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 모바일 디바이스(1200)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 모바일 디바이스(1200)에 설치되거나 또는 모바일 디바이스(1200)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0098] [0113] 도 13은 도 1의 디바이스(100)의 양상들을 통합하는 항공 디바이스(1300)를 예시한다. 도 13에서, 항공 디바이스(1300)는 도 1의 디바이스(100)를 포함하거나 그에 결합된다. 항공 디바이스(1300)는 유인, 무인 또는 원격 조종 항공 디바이스(예컨대, 패키지 배달 드론)이다. 항공 디바이스(1300)는 제어 시스템(1302) 및 센서들(134), 이를테면, 도 1의 마이크로폰(들)(104), 도 6의 카메라(들)(650), 도 6의 포지션 센서(들)(652), 또는 도 6의 다른 센서(들)(654)를 포함한다. 제어 시스템(1302)은 화물 반출, 센서 활성화, 이륙, 내비게이션, 착륙 또는 이들의 조합들과 같은 항공 디바이스(1300)의 다양한 동작들을 제어한다. 예컨대, 제어 시스템(1302)은 지정된 지점들 사이의 항공 디바이스(1300)의 비행 및 특정 위치에서의 화물 전개를 제어할 수 있다. 특정 양상에서, 제어 시스템(1302)은 디바이스(100)에 의한 특정 사운드 이벤트의 검출에 응답하여 하나 이상의 액션을 수행한다. 예시하자면, 제어 시스템(1302)은 디바이스(100)가 항공기 엔진을 검출하는 것에 대한 응답으로 안전 착륙 프로토콜을 개시할 수 있다.

[0099] [0114] 도 13에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 항공 디바이스(1300)에 설치되거나 모바일 디바이스(1300)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 항공 디바이스(1300)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 항공 디바이스(1300)에 설치되거나 또는 모바일 디바이스(1300)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0100] [0115] 도 14는 도 1의 디바이스(100)의 양상들을 통합하는 헤드셋(1400)을 예시한다. 도 14에서, 헤드셋(1400)은 도 1의 디바이스(100)를 포함하거나 그에 결합된다. 헤드셋(1400)은 사용자의 스피치를 주로 캡처하도록 포지셔닝된 도 1의 마이크로폰(들)(104)을 포함한다. 헤드셋(1400)은 또한 (예컨대, 잡음 제거 동작들을 위해) 환경 사운드들을 주로 캡처하도록 포지셔닝된 하나 이상의 부가적인 마이크로폰 및 카메라(들)(650), 포지션 센서(들)(652), 또는 도 6의 다른 센서(들)(654)와 같은 센서들(134) 중 하나 이상을 포함할 수 있다. 특정 양상에서, 헤드셋(1400)은 디바이스(100)에 의한 특정 사운드 이벤트의 검출에 응답하여 하나 이상의 액션들을 수행한다. 예시하자면, 헤드셋(1400)은 디바이스(100)가 총성을 검출한 것에 대한 응답으로 잡음 소거 피처를 활성화시킬 수 있다. 헤드셋(1400)은 또한 SEC 모델들(114) 중 하나 이상을 업데이트할 수 있다.

[0101] [0116] 도 14에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 헤드셋(1400)에 설치되거나 헤드셋(1400)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 헤드셋(1400)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 헤드셋(1400)에 설치되거나 또는 헤드셋(1400)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0102] [0117] 도 15는 도 1의 디바이스(100)의 양상들을 통합하는 기기(1500)를 예시한다. 도 15에서, 기기(1500)는

램프이지만; 다른 구현들에서, 기기(1500)는 다른 사물 인터넷(Internet-of-Things) 기기, 이를테면, 냉장고, 커피 메이커, 오븐, 다른 가전제품 등을 포함한다. 기기(1500)는 도 1의 디바이스(100)를 포함하거나 그에 결합된다. 기기(1500)는 센서들(134), 이를테면, 도 1의 마이크로폰(들)(104), 도 6의 카메라(들)(650), 도 6의 포지션 센서(들)(652), 또는 도 6의 다른 센서(들)(654)를 포함한다. 특정 양상에서, 기기(1500)는 디바이스(100)에 의한 특정 사운드 이벤트의 검출에 응답하여 하나 이상의 액션들을 수행한다. 예시하자면, 기기(1500)는 디바이스(100)가 문 닫힘을 검출한 것에 대한 응답으로 조명을 활성화시킬 수 있다. 기기(1500)는 또한 SEC 모델들(114) 중 하나 이상을 업데이트할 수 있다.

[0103] [0118] 도 15에 예시된 예에서, 디바이스(100)는 SEC 모델들(114), SEC 엔진(120), 드리프트 검출기(128), 장면 검출기(140) 및 모델 업데이터(152)를 포함한다. 다른 구현들에서, 디바이스(100)는 기기(1500)에 설치되거나 헤드셋(1500)에서 사용될 때, 모델 업데이터(152)를 생략한다. 예시하자면, 도 6의 모델 업데이트 데이터(148)는 원격 컴퓨팅 디바이스(618)에 전송될 수 있고, 원격 컴퓨팅 디바이스(618)는 모델 업데이트 데이터(148)에 기반하여 SEC 모델들(114) 중 하나를 업데이트할 수 있다. 이러한 구현들에서, 업데이트된 SEC 모델(114)은 SEC 엔진(120)에 의해 사용되도록 기기(1500)에 다운로드될 수 있다. 일부 구현들에서, 디바이스(100)는, 기기(1500)에 설치되거나 또는 기기(1500)에서 사용될 때, 도 3의 모델 체커(320)를 더 포함한다.

[0104] [0119] 도 16은 도 1의 디바이스(100)의 동작 방법(1600)의 예를 예시하는 흐름도이다. 방법(1600)은 디바이스(100)에 의해 개시, 제어 또는 수행될 수 있다. 예컨대, 도 6의 프로세서(들)(604 또는 606)는 메모리(608)로부터 명령들(660)을 실행하여 드리프트 검출기(128)가 모델 업데이트 데이터(148)를 생성하게 할 수 있다.

[0105] [0120] 블록(1602)에서, 방법(1600)은 오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하는 단계를 포함한다. 예컨대, 도 1의 SEC 엔진(120)(또는 SEC 엔진(120)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))은 오디오 데이터 샘플들(110)을 입력으로서 SEC 모델(112)에 제공할 수 있다. 일부 구현들에서, 방법(1600)은 또한 오디오 데이터 샘플들에 대응하는 오디오 데이터를 캡처하는 단계를 포함한다. 예컨대, 도 1의 마이크로폰(들)(104)은 마이크로폰(들)(104)에 의해 검출된 사운드(102)에 기반하여 오디오 신호들(106)을 생성할 수 있다. 추가로, 일부 구현들에서, 방법(1600)은 메모리에 저장된 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 포함한다. 예컨대, 도 1의 SEC 엔진(120)(또는 SEC 엔진(120)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))은, 오디오 데이터 샘플들과 연관되는 센서 데이터에 기반하여, 오디오 장면 또는 SEC 모델(112)을 식별하는 입력에 기반하여, 오디오 데이터 샘플들이 수신되는 경우에 기반하여, 설정 데이터에 기반하여, 또는 이들의 조합에 기반하여, 이용 가능한 SEC 모델들(114) 중에서 SEC 모델(112)을 선택할 수 있다.

[0106] [0121] 블록(1604)에서, 방법(1600)은, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하는 단계를 포함한다. 예컨대, 도 1의 SEC 엔진(120)(또는 SEC 엔진(120)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))은 오디오 데이터 샘플들(110)의 사운드 클래스(122)가 SEC 모델(112)에 의해 인식되었는지 여부를 결정할 수 있다. 예시하자면, SEC 모델(112)은 SEC 모델(112)이 트레이닝되어 인식할 각각의 사운드 클래스와 연관된 신뢰도 메트릭을 생성할 수 있고, 사운드 클래스가 SEC 모델에 의해 인식되었는지 여부의 결정은 신뢰도 메트릭(들)의 값들에 기초할 수 있다. 특정 양상에서, 사운드 클래스(122)가 SEC 모델(112)에 의해 인식되었다는 결정에 기반하여, SEC 엔진(120)은 오디오 데이터 샘플들(110)과 연관된 사운드 클래스(122)를 표시하는 출력(124)을 생성한다.

[0107] [0122] 블록(1606)에서, 방법(1600)은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하는 단계를 포함한다. 예컨대, 도 1의 드리프트 검출기(128)(또는 드리프트 검출기(128)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))는 SEC 모델(112)이 오디오 데이터 샘플들(110)과 연관된 오디오 장면(142)에 대응하는지 여부를 결정할 수 있다. 예시하자면, 장면 검출기(140)는, 입력 디바이스(136)를 통해 수신된 입력에 기반하여, 센서(들)(134)로부터의 센서 데이터에 기반하여, 클록(132)에 의해 표시된 바와 같은 오디오 데이터 샘플들(110)과 연관된 타임스탬프에 기반하여, 설정 데이터(130)에 기반하여, 또는 이들의 조합에 기반하여, 오디오 장면(142)을 결정할 수 있다. 이용 가능한 SEC 모델들 중에서 SEC 모델(112)이 선택되는 구현들에서, 오디오 장면(142)의 결정은 SEC 모델(112)을 선택하는 데 사용된 정보와 상이한 정보에 기반할 수 있다.

[0108] [0123] 블록(1608)에서, 방법(1600)은 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하는 단계를 포함

한다. 예컨대, 도 1의 드리프트 검출기(128)(또는 드리프트 검출기(128)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))는 오디오 데이터 샘플들(110)에 기반하여 모델 업데이트 데이터(148)를 저장할 수 있다. 특정 양상에서, SEC 모델(112)이 오디오 데이터 샘플들(110)과 연관된 오디오 장면(142)에 대응한다고 드리프트 검출기(128)가 결정하면, 드리프트 검출기(128)는 모델 업데이트 데이터(148)로서 드리프트 데이터(144)를 저장하고, 그리고 드리프트 검출기(128)가 SEC 모델(112)이 오디오 데이터 샘플들(110)과 연관된 오디오 장면(142)에 대응하지 않는다고 결정하면, 드리프트 검출기(128)는 미지의 데이터(146)를 모델 업데이트 데이터(148)로서 저장한다.

[0109] [0124] 방법(1600)은 또한 모델 업데이트 데이터에 기반하여 SEC 모델을 업데이트하는 단계를 포함할 수 있다. 예컨대, 도 1의 모델 업데이터(152)(또는 모델 업데이터(152)에 대응하는 명령들(660)을 실행하는 도 6의 프로세서(들)(604, 606))은 도 2를 참조하여 설명된 바와 같은, 도 3을 참조하여 설명된 바와 같은, 또는 둘 모두와 같은 모델 업데이트 데이터(148)에 기반하여 SEC 모델(112)을 업데이트할 수 있다.

[0110] [0125] 특정 양상에서, 방법(1600)은 모델 업데이트 데이터를 저장한 후, 모델 업데이트 데이터의 임계 양(quantity amount)이 누적되었는지 여부를 결정하는 단계를 포함한다. 예컨대, 모델 업데이터(152)는, 도 1의 모델 업데이트 데이터(148)가 SEC 모델(112)의 업데이트 트레이닝을 개시하기에 충분한 데이터(예컨대, 특정 SEC 모델 및 특정 사운드 클래스와 연관된 모델 업데이트 데이터(148)의 임계 수량)를 포함하는 경우를 결정할 수 있다. 방법(1600)은 또한, 모델 업데이트 데이터의 임계량이 누적되었다는 결정에 기반하여, 누적된 모델 업데이트 데이터를 사용하여 사운드 이벤트 분류 모델의 자동 업데이트를 개시하는 단계를 포함할 수 있다. 예컨대, 모델 업데이터(152)는 사용자로부터의 입력 없이 모델 업데이트를 개시할 수 있다. 특정 구현에서, 자동 업데이트는 SEC 모델(112)을 미세 튜닝하여 업데이트된 SEC 모델(154)을 생성한다. 예컨대, 자동 업데이트 전에, SEC 모델(112)은 특정 사운드 클래스의 다수의 변형들을 인식하도록 트레이닝되었으며, 자동 업데이트는 SEC 모델(112)을 수정하여 SEC 모델(112)이 특정 사운드 클래스의 추가 변형을 특정 사운드 클래스에 대응하는 것으로 인식하게 한다.

[0111] [0126] 트레이닝이 발생하는 방식으로 인해, SEC 모델은 일반적으로 폐쇄형-세트(closed-set)이다. 즉, SEC 모델이 인식할 수 있는 사운드 클래스들의 수 및 타입은 트레이닝 동안 고정되고 제한된다. 트레이닝 후에, SEC 모델은 통상적으로 입력과 출력 사이에 정적 관계를 갖는다. 입력과 출력 사이의 이러한 정적 관계는, 트레이닝 동안 학습된 맵핑이 미래에(예컨대, 새로운 데이터를 평가할 때) 유효하고, 입력 및 출력 데이터 사이의 관계들이 변경되지 않음을 의미한다. 그러나, 각각의 사운드 클래스에 대한 트레이닝 샘플들의 포괄적인 세트를 수집하는 것은 어렵고, 포괄적이고 정교한 SEC 모델을 트레이닝하기 위해 이용 가능한 트레이닝 데이터 전부에 적절히 주석을 다는 것은 어렵다.

[0112] [0127] 대조적으로, 사용 동안, SEC 모델들은 개방-세트(open-set) 문제에 직면한다. 예컨대, 사용 동안, SEC 모델에는 알려진 사운드 이벤트 및 미지의 사운드 이벤트들 둘 모두와 연관된 데이터 샘플들이 제공될 수 있다. 추가적으로, SEC 모델이 트레이닝되어 인식할 각각의 사운드 클래스에서의 사운드들 또는 사운드 특징들의 분포는 시간이 지남에 따라 변할 수 있거나, 또는 이용 가능한 트레이닝 데이터에서 포괄적으로 표현되지 않을 수 있다. 예컨대, 교통 사운드들의 경우, 위치들, 시간들, 혼잡한 또는 혼잡하지 않은 교차로 등에 기반한 사운드들의 차이들은 교통 사운드 클래스에 대한 트레이닝 데이터에서 명시적으로 캡처되지 않을 수 있다. 이러한 이유들 및 다른 이유들로 인해, SEC 모델을 훈련시키는 데 사용되는 훈련 데이터와 사용 중에 SEC 모델이 제공되는 데이터세트 사이에 불일치들이 있을 수 있다. 이러한 불일치들(예컨대, 데이터세트 시프트 또는 드리프트)은 위치, 시간, 사운드 신호를 캡처하는 디바이스 등과 같은 다양한 팩터들에 의존한다. 데이터세트 시프트는 SEC 모델로부터 불량한 예측 결과들을 초래할 수 있다. 개시된 시스템들 및 방법들은 감독이 거의 또는 전혀 없이 이러한 시프트 데이터를 검출하도록 SED 모델을 적응시킴으로써 이들 및 다른 문제들을 극복한다. 추가적으로, 일부 양상들에서, SEC 모델은 이전의 트레이닝된 사운드 클래스들을 잊어버리지 않으면서 새로운 사운드 클래스들을 인식하도록 업데이트될 수 있다.

[0113] [0128] 특정 양상에서, 시스템이 추론 모드에서 동작하고 있는 동안에는 SEC 모델들의 어떠한 트레이닝도 수행되지 않는다. 오히려, 추론 모드에서의 동작 동안, 검출된 사운드들을 분석하기 위해, 하나 이상의 이전에 트레이닝된 SEC 모델들의 형태의 기존 지식이 사용된다. 하나 초과의 SEC 모델이 사운드를 분석하는 데 사용될 수 있다. 예컨대, SEC 모델들의 앙상블이 추론 모드에서의 동작 동안 사용될 수 있다. 특정 SEC는 트리거 조건의 검출에 기반하여 한 세트의 이용 가능한 SEC 모델들로부터 선택될 수 있다. 예시하자면, 특정 트리거(또는 트리거들)가 활성화될 때마다, 활성 SEC 모델로서 또한 "소스 SEC 모델"로 지칭될 수 있는 특정 SEC 모델이 사용될 것이다. 트리거(들)는 로케이션들, 사운드들, 카메라 정보, 다른 센서 데이터, 사용자 입력 등에 기반할 수 있

다. 예컨대, 특정 SEC 모델은, 혼잡한 영역들, 이를테면, 테마 파크들, 실외 쇼핑몰들, 공공 광장 등과 관련된 사운드 이벤트들을 인식하도록 트레이닝될 수 있다. 이러한 예에서, 사운드를 캡처하는 디바이스가 이들 로케이션들 중 임의의 로케이션에 있음을 글로벌 포지셔닝 데이터가 표시할 때, 특정 SEC 모델은 활성 SEC 모델로서 사용될 수 있다. 이러한 예에서, 트리거는 사운드를 캡처하는 디바이스의 로케이션에 기반하며, 디바이스가 로케이션에 있는 것으로 검출될 때, 활성 SEC 모델이 (예컨대, 이전의 활성 SEC 모델에 부가하여 또는 그 대신에) 선택되고 로딩된다.

[0114] [0129] 설명된 구현들과 관련하여, 장치는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단을 포함한다. 예컨대, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단은, 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), SEC 엔진(120), 특징 추출기(108), 마이크로폰(들)(104), 코덱(624), 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0115] [0130] 장치는 또한, 사운드 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단을 포함한다. 예컨대, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단은 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), SEC 엔진(120), 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0116] [0131] 장치는 또한, 사운드 클래스가 인식되지 않았다는 결정에 응답하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단을 포함한다. 예컨대, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단은, 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), 드리프트 검출기(128), 장면 검출기(140), 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0117] [0132] 장치는 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 응답하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하기 위한 수단을 포함한다. 예컨대, 모델 업데이트 데이터를 저장하기 위한 수단은, 원격 컴퓨팅 디바이스(618), 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), 드리프트 검출기(128), 메모리(608), 모델 업데이트 데이터를 저장하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0118] [0133] 일부 구현들에서, 장치는 선택 기준에 기반하여 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하기 위한 수단을 포함한다. 예컨대, 사운드 이벤트 분류 모델을 선택하기 위한 수단은, 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), SEC 엔진(120), 사운드 이벤트 분류 모델을 선택하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0119] [0134] 일부 구현들에서, 장치는 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하기 위한 수단을 포함한다. 예컨대, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하기 위한 수단은, 원격 컴퓨팅 디바이스(618), 디바이스(100), 명령들(660), 프로세서(604), 프로세서(들)(606), 모델 업데이터(152), 모델 체커(320), 사운드 이벤트 분류 모델을 업데이트하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합을 포함한다.

[0120] [0135] 당업자들은, 본원에 개시된 구현들과 관련하여 설명된 다양한 예시적인 로직 블록들, 구성들, 모듈들, 회로들, 및 알고리즘 단계들이 전자 하드웨어, 프로세서에 의해 실행되는 컴퓨터 소프트웨어, 또는 이 둘의 조합들로서 구현될 수 있음을 추가로 인식할 것이다. 다양한 예시적인 컴포넌트들, 블록들, 구성들, 모듈들, 회로들, 및 단계들은 일반적으로 그들의 기능성의 관점에서 위에서 설명되었다. 이러한 기능이 하드웨어로 구현되는지 아니면 프로세서 실행 가능 명령들로 구현되는지는 특정 애플리케이션 및 전체 시스템에 대해 부과된 설계 제한들에 의존한다. 당업자들은 설명된 기능을 각각의 특정 애플리케이션에 대해 다양한 방식들로 구현할 수 있으며, 이러한 구현 결정들이 본 개시내용의 범위로부터 벗어나는 것으로 해석되지 않아야 한다.

[0121] [0136] 본 명세서에 개시된 구현들과 관련하여 설명된 방법 또는 알고리즘의 단계들은 직접 하드웨어로, 프로세서에 의해 실행되는 소프트웨어 모듈로, 또는 이 둘의 결합으로 구현될 수 있다. 소프트웨어 모듈은, 랜덤 액세스 메모리(RAM), 플래시 메모리, 판독-전용 메모리(ROM), 프로그램가능 판독-전용 메모리(PROM), 소거가능 프로

그램가능 판독-전용 메모리(EPROM), 전기적으로 소거 가능한 프로그램가능 판독-전용 메모리(EEPROM), 레지스터들, 하드 디스크, 제거가능 디스크, 콤팩트 디스크 판독-전용 메모리(CD-ROM; compact disc read-only memory), 또는 당해 기술분야에 알려진 임의의 다른 형태의 비일시적 저장 매체에 상주할 수 있다. 예시적인 저장 매체는, 프로세서가 저장 매체로부터 정보를 판독하고, 저장 매체에 정보를 기록할 수 있도록 프로세서에 커플링된다. 대안으로, 저장 매체는 프로세서에 통합될 수 있다. 프로세서 및 저장 매체는 주문형 집적 회로(ASIC)에 상주할 수 있다. ASIC는 컴퓨팅 디바이스 또는 사용자 단말에 상주할 수 있다. 대안적으로, 프로세서 및 저장 매체는 컴퓨팅 디바이스 또는 사용자 단말에서 개별 컴포넌트들로서 상주할 수 있다.

[0122] [0137] 본 개시내용의 특정 양상들은 아래의 제1 세트의 상호관련된 조항들에서 설명된다:

[0123] [0138] 조항 1에 따르면, 디바이스는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 추가로 구성된다. 하나 이상의 프로세서들은 또한, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하도록 구성된다. 하나 이상의 프로세서들은 추가로, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하도록 구성된다.

[0124] [0139] 조항 2는 조항 1의 디바이스를 포함하고, 하나 이상의 프로세서들에 커플링되고 오디오 데이터 샘플들에 대응하는 오디오 데이터를 캡처하도록 구성된 마이크로폰을 더 포함한다.

[0125] [0140] 조항 3은 조항 1 또는 2의 디바이스를 포함하고, 하나 이상의 프로세서들에 커플링되고 복수의 사운드 이벤트 분류 모델들을 저장하도록 구성된 메모리를 더 포함하고, 하나 이상의 프로세서들은 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하도록 구성된다.

[0126] [0141] 조항 4는 조항 3의 디바이스를 포함하고, 오디오 데이터 샘플들과 연관된 센서 데이터를 생성하도록 구성된 하나 이상의 센서들을 더 포함하고, 하나 이상의 프로세서들은 센서 데이터에 기반하여 사운드 이벤트 분류 모델을 선택하도록 구성된다.

[0127] [0142] 조항 5는 조항 4의 디바이스를 포함하고, 하나 이상의 센서들은 카메라 및 포지션 센서를 포함한다.

[0128] [0143] 조항 6은 조항 3 내지 5중 어느 한 조항의 디바이스를 포함하고, 오디오 장면을 식별하는 입력을 수신하도록 구성되는 하나 이상의 입력 디바이스들을 더 포함하고, 하나 이상의 프로세서들은 오디오 장면에 기반하여 사운드 이벤트 분류 모델을 선택하도록 구성된다.

[0129] [0144] 조항 7은 조항 3 내지 6 중 어느 한 조항의 디바이스를 포함하고, 하나 이상의 프로세서들은, 오디오 데이터 샘플들이 수신되는 경우에 기반하여 사운드 이벤트 분류 모델을 선택하도록 구성된다.

[0130] [0145] 조항 8은 조항 3 내지 8 중 어느 한 조항의 디바이스를 포함하고, 메모리는 하나 이상의 디바이스 설정들을 나타내는 설정 데이터를 추가로 저장하고, 그리고 하나 이상의 프로세서들은 설정 데이터에 기반하여 사운드 이벤트 분류 모델을 선택하도록 구성된다.

[0131] [0146] 조항 9는 조항 1 내지 8 중 어느 한 조항의 디바이스를 포함하고, 하나 이상의 프로세서들은, 사운드 클래스가 인식되었다는 결정에 기반하여, 오디오 데이터 샘플들과 연관된 사운드 클래스를 표시하는 출력을 생성하도록 추가로 구성된다.

[0132] [0147] 조항 10은 조항 1 내지 9 중 어느 한 조항의 디바이스를 포함하고, 하나 이상의 프로세서들은, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하지 않는다는 결정에 기반하여, 오디오 데이터 샘플들에 대응하는 오디오 데이터를 새로운 사운드 이벤트 분류 모델에 대한 트레이닝 데이터로서 저장하도록 추가로 구성된다.

[0133] [0148] 조항 11은 조항 1 내지 10 중 어느 한 조항의 디바이스를 포함하고, 사운드 이벤트 분류 모델은 출력과 연관된 신뢰도 메트릭을 생성하도록 추가로 구성되고, 그리고 하나 이상의 프로세서들은, 사운드 클래스가 신뢰도 메트릭에 기반하여 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 구성된다.

[0134] [0149] 조항 12는 조항 1 내지 11 중 어느 한 조항의 디바이스를 포함하고, 하나 이상의 프로세서들은 추가로, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하도록 구성된다.

[0135] [0150] 조항 13은 조항 1 내지 12 중 어느 한 조항의 디바이스를 포함하고, 오디오 장면을 식별하는 입력을 수신하도록 구성되는 하나 이상의 입력 디바이스들을 더 포함하고, 하나 이상의 프로세서들은 사운드 이벤트 분류 모델이 입력에 기반하여 오디오 장면에 대응하는지 여부를 결정하도록 구성된다.

[0136] [0151] 조항 14는 조항 1 내지 13 중 어느 한 조항의 디바이스를 포함하고, 오디오 데이터 샘플들과 연관된 센서 데이터를 생성하도록 구성된 하나 이상의 센서들을 더 포함하고, 하나 이상의 프로세서들은 사운드 이벤트 분류 모델이 센서 데이터에 기반하여 오디오 장면에 대응하는지 여부를 결정하도록 구성된다.

[0137] [0152] 조항 15는 조항 14의 디바이스를 포함하고, 하나 이상의 센서들은 카메라 및 포지션 센서를 포함한다.

[0138] [0153] 조항 16은 조항 14 또는 15의 디바이스를 포함하고, 하나 이상의 프로세서들은 추가로, 오디오 데이터 샘플들과 연관된 타임스탬프에 기반하여, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부를 결정하도록 구성된다.

[0139] [0154] 조항 17은 조항 1 내지 16 중 어느 한 조항의 디바이스를 포함하며, 하나 이상의 프로세서들은 모바일 컴퓨팅 디바이스 내에 통합된다.

[0140] [0155] 조항 18은 조항 1 내지 16 중 어느 한 조항의 디바이스를 포함하며, 하나 이상의 프로세서들은 차량 내에 통합된다.

[0141] [0156] 조항 19는 조항 1 내지 16 중 어느 한 조항의 디바이스를 포함하며, 하나 이상의 프로세서들은 웨어러블 디바이스 내에 통합된다.

[0142] [0157] 조항 20은 조항 1 내지 16 중 어느 한 조항의 디바이스를 포함하며, 하나 이상의 프로세서들은 증강 현실 헤드셋, 혼합 현실 헤드셋 또는 가상 현실 헤드셋 내에 통합된다.

[0143] [0158] 조항 21은 조항 1 내지 20 중 어느 한 조항의 디바이스를 포함하며, 하나 이상의 프로세서들은 집적 회로에 포함된다.

[0144] [0159] 조항 22는 조항 1 내지 21 중 어느 한 조항의 디바이스를 포함하며, 사운드 이벤트 분류 모델은 특정 사운드 클래스를 인식하도록 트레이닝되고, 모델 업데이트 데이터는, 사운드 이벤트 분류 모델이 특정 사운드 클래스에 대응하는 것으로 인식하도록 트레이닝되지 않은, 특정 사운드 클래스 내의 사운드의 특성들의 변동을 표현하는 드리프트 데이터를 포함한다.

[0145] [0160] 본 개시내용의 특정 양상들은 아래의 제2 세트의 상호관련된 조항들에서 설명된다:

[0146] [0161] 조항 23에 따르면, 방법은 하나 이상의 프로세서들이, 오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하는 단계를 포함한다. 방법은 또한, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 하나 이상의 프로세서들이 결정하는 단계를 포함한다. 방법은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 하나 이상의 프로세서들이 결정하는 단계를 더 포함한다. 방법은 또한, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 하나 이상의 프로세서들이 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하는 단계를 포함한다.

[0147] [0162] 조항 24는 조항 23의 방법을 포함하고, 하나 이상의 프로세서들에 커플링된 메모리에 저장된 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0148] [0163] 조항 25는 조항 24의 방법을 포함하고, 사운드 이벤트 분류 모델은 사용자 입력, 설정 데이터, 로케이션 데이터, 이미지 데이터, 비디오 데이터, 오디오 데이터 샘플들과 연관된 타임스탬프 또는 이들의 조합에 기반하여 선택된다.

[0149] [0164] 조항 26은 조항 23 내지 25 중 어느 한 조항의 방법을 포함하고, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은, 사운드 이벤트 분류 모델에 의해 생성된 신뢰도 메트릭, 사용자 입력, 설정 데이터, 로케이션 데이터, 이미지 데이터, 비디오 데이터, 오디오 데이터 샘플들과 연관된 타임스탬프 또는 이들의 조합에 기반한다.

[0150] [0165] 조항 27은 조항 23 내지 26 중 어느 한 조항의 방법을 포함하며, 오디오 데이터 샘플들에 대응하는 오디오 데이터를 캡처하는 단계를 더 포함한다.

[0151] [0166] 조항 28은 조항 23 내지 27 중 어느 한 조항의 방법을 포함하며, 복수의 이용 가능한 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0152] [0167] 조항 29는 조항 23 내지 27 중 어느 한 조항의 방법을 포함하며, 오디오 데이터 샘플들과 연관된 센서 데이터를 수신하는 단계, 및 센서 데이터에 기반하여 복수의 이용 가능한 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0153] [0168] 조항 30은 조항 23 내지 27 중 어느 한 조항의 방법을 포함하고, 오디오 장면을 식별하는 입력을 수신하는 단계, 및 오디오 장면에 기반하여 복수의 이용 가능한 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0154] [0169] 조항 31은 조항 23 내지 27 중 어느 한 조항의 방법을 포함하며, 오디오 데이터 샘플들이 수신되는 경우에 기반하여 복수의 이용 가능한 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0155] [0170] 조항 32는 조항 23 내지 27 중 어느 한 조항의 방법을 포함하며, 설정 데이터에 기반하여 복수의 이용 가능한 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하는 단계를 더 포함한다.

[0156] [0171] 조항 33은 조항 23 내지 32 중 어느 한 조항의 방법을 포함하며, 사운드 클래스가 인식되었다는 결정에 기반하여, 오디오 데이터 샘플들과 연관된 사운드 클래스를 표시하는 출력을 생성하는 단계를 더 포함한다.

[0157] [0172] 조항 34는 조항 23 내지 33 중 어느 한 조항의 방법을 포함하며, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하지 않는다는 결정에 기반하여, 오디오 데이터 샘플들에 대응하는 오디오 데이터를 새로운 사운드 이벤트 분류 모델에 대한 트레이닝 데이터로서 저장하는 단계를 더 포함한다.

[0158] [0173] 조항 35는, 조항 23 내지 34 중 어느 한 조항의 방법을 포함하며, 사운드 이벤트 분류 모델의 출력은 신뢰도 메트릭을 포함하고, 방법은 신뢰도 메트릭에 기반하여 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하는 단계를 더 포함한다.

[0159] [0174] 조항 36은 조항 23 내지 35 중 어느 한 조항의 방법을 포함하며, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하는 단계를 더 포함한다.

[0160] [0175] 조항 37은 조항 23 내지 36 중 어느 한 조항의 방법을 포함하며, 오디오 장면을 식별하는 입력을 수신하는 단계를 더 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 입력에 기반한다.

[0161] [0176] 조항 38은 조항 23 내지 37 중 어느 한 조항의 방법을 포함하며, 오디오 데이터 샘플들과 연관된 센서 데이터를 수신하는 단계를 더 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 센서 데이터에 기반한다.

[0162] [0177] 조항 39는 조항 23 내지 38 중 어느 한 조항의 방법을 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부에 대한 결정은 오디오 데이터 샘플들과 연관된 타임스탬프에 기반한다.

[0163] [0178] 조항 40은 조항 23 내지 39 중 어느 한 조항의 방법을 포함하며, 모델 업데이트 데이터를 저장한 후에, 모델 업데이트 데이터의 임계량이 누적되었는지 여부를 결정하는 단계, 및 모델 업데이트 데이터의 임계량이 누적되었다는 결정에 기반하여, 누적된 모델 업데이트 데이터를 사용하여 사운드 이벤트 분류 모델의 자동 업데이트를 개시하는 단계를 더 포함한다.

[0164] [0179] 조항 41은 조항 23 내지 40 중 어느 한 조항의 방법을 포함하며, 자동 업데이트 전에, 사운드 이벤트 분류 모델은 특정 사운드 클래스의 다수의 변화들을 인식하도록 트레이닝되었고, 그리고 자동 업데이트는 사운드 이벤트 분류 모델이 특정 사운드 클래스에 대응하는 것으로 특정 사운드 클래스의 추가적인 변형을 인식할 수 있게 하도록 사운드 이벤트 분류 모델을 수정한다.

[0165] [0180] 본 개시내용의 특정 양상들은 아래의 제3 세트의 상호관련된 조항들에서 설명된다:

[0166] [0181] 조항 42에 따르면, 디바이스는 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단을 포함한다. 디바이스는 또한, 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단을 포함한다. 디바이스는, 사운드 클래스가 인식되지 않았다는 결정에 응답하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단을 더 포함한다. 디바이스는 또한, 사운드 이벤트 분류

모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 응답하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하기 위한 수단을 포함한다.

[0167] [0182] 조항 43은 조항 42의 디바이스를 포함하며, 선택 기준에 기반하여 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하기 위한 수단을 더 포함한다.

[0168] [0183] 조항 44는 조항 42 또는 43의 디바이스를 포함하고, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하기 위한 수단을 더 포함한다.

[0169] [0184] 조항 45는 조항 42 내지 44 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들에 대응하는 오디오 데이터를 캡처하기 위한 수단을 더 포함한다.

[0170] [0185] 조항 46은 조항 42 내지 45 중 어느 한 조항의 디바이스를 포함하며, 복수의 사운드 이벤트 분류 모델들을 저장하기 위한 수단 및 복수의 사운드 이벤트 분류 모델들로부터 사운드 이벤트 분류 모델을 선택하기 위한 수단을 더 포함한다.

[0171] [0186] 조항 47은 조항 46의 디바이스를 포함하며, 오디오 장면을 식별하는 입력을 수신하기 위한 수단을 더 포함하며, 사운드 이벤트 분류 모델은 오디오 장면을 식별하는 입력에 기반하여 선택된다.

[0172] [0187] 조항 48은 조항 46의 디바이스를 포함하며, 오디오 데이터 샘플들이 수신된 때를 결정하기 위한 수단을 더 포함하며, 사운드 이벤트 분류 모델은 오디오 데이터 샘플들이 수신된 때를 기초로 선택된다.

[0173] [0188] 조항 49는 조항 46의 디바이스를 포함하며, 하나 이상의 디바이스 설정들을 나타내는 설정 데이터를 저장하기 위한 수단을 더 포함하며, 사운드 이벤트 분류 모델은 설정 데이터에 기반하여 선택된다.

[0174] [0189] 조항 50은 조항 42 내지 49 중 어느 한 조항의 디바이스를 포함하고, 사운드 클래스가 인식되었다는 결정에 기반하여 오디오 데이터 샘플들과 연관된 사운드 클래스를 표시하는 출력을 생성하기 위한 수단을 더 포함한다.

[0175] [0190] 조항 51은 조항 42 내지 50 중 어느 한 조항의 디바이스를 포함하고, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하지 않는다는 결정에 기반하여, 새로운 사운드 이벤트 분류 모델에 대한 트레이닝 데이터로서 오디오 데이터 샘플들에 대응하는 오디오 데이터를 저장하기 위한 수단을 더 포함한다.

[0176] [0191] 조항 52는 조항 42 내지 51 중 어느 한 조항의 디바이스를 포함하고, 사운드 이벤트 분류 모델은 출력과 연관된 신뢰도 메트릭을 생성하도록 추가로 구성되고, 그리고 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부의 결정은 신뢰도 메트릭에 기반한다.

[0177] [0192] 조항 53은 조항 42 내지 52의 디바이스를 포함하고, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하기 위한 수단을 더 포함한다.

[0178] [0193] 조항 54는 조항 42 내지 53 중 어느 한 조항의 디바이스를 포함하며, 오디오 장면을 식별하는 입력을 수신하기 위한 단계를 더 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 입력에 기반한다.

[0179] [0194] 조항 55는 조항 42 내지 54 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들과 연관된 센서 데이터를 수신하기 위한 수단을 더 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 센서 데이터에 기반한다.

[0180] [0195] 조항 56은 조항 42 내지 55 중 어느 한 조항의 디바이스를 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부에 대한 결정은 오디오 데이터 샘플들과 연관된 타임스탬프에 기반한다.

[0181] [0196] 조항 57은 조항 42 내지 56 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단, 사운드 이벤트 분류 모델의 출력을 수신하기 위한 수단, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단, 및 오디오 데이터 샘플들에 기반하여 모델 업데이트 데이터를 저장하기 위한 수단은 모바일 컴퓨팅 디바이스 내에 통합된다.

[0182] [0197] 조항 58은 조항 42 내지 56 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들을 사운드 이

벤트 분류 모델에 제공하기 위한 수단, 사운드 이벤트 분류 모델의 출력을 수신하기 위한 수단, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단, 및 오디오 데이터 샘플들에 기반하여 모델 업데이트 데이터를 저장하기 위한 수단은 차량 내에 통합된다.

[0183] [0198] 조항 59는 조항 42 내지 56 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단, 사운드 이벤트 분류 모델의 출력을 수신하기 위한 수단, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단, 및 오디오 데이터 샘플들에 기반하여 모델 업데이트 데이터를 저장하기 위한 수단은 웨어러블 디바이스 내에 통합된다.

[0184] [0199] 조항 60은 조항 42 내지 56 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단, 사운드 이벤트 분류 모델의 출력을 수신하기 위한 수단, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단, 및 오디오 데이터 샘플들에 기반하여 모델 업데이트 데이터를 저장하기 위한 수단은 증강 현실 헤드셋, 혼합 현실 헤드셋, 또는 가상 현실 헤드셋 내에 통합된다.

[0185] [0200] 조항 61은 조항 42 내지 60 중 어느 한 조항의 디바이스를 포함하며, 오디오 데이터 샘플들을 사운드 이벤트 분류 모델에 제공하기 위한 수단, 사운드 이벤트 분류 모델의 출력을 수신하기 위한 수단, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하기 위한 수단, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하기 위한 수단, 및 오디오 데이터 샘플들에 기반하여 모델 업데이트 데이터를 저장하기 위한 수단은 집적 회로 내에 포함된다.

[0186] [0201] 본 개시내용의 특정 양상들은 아래의 제4 세트의 상호관련된 조항들에서 설명된다:

[0187] [0202] 조항 62에 따르면, 비일시적 컴퓨터 판독 가능 매체는, 프로세서로 하여금 오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공하게 하는, 프로세서에 의해 실행 가능한 명령들을 저장한다. 명령들은 또한, 오디오 데이터 샘플들에 대한 응답으로 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정하도록 프로세서에 의해 실행 가능하다. 명령들은, 사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정하도록 프로세서에 의해 추가로 실행 가능하다. 명령들은, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장하도록 프로세서에 의해 추가로 실행 가능하다.

[0188] [0203] 조항 63은 조항 62의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 추가로, 프로세서로 하여금, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하게 한다.

[0189] [0204] 조항 64는 조항 62 또는 63의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 추가로 프로세서로 하여금, 메모리에 저장된 복수의 사운드 이벤트 분류 모델들 중에서 사운드 이벤트 분류 모델을 선택하게 한다.

[0190] [0205] 조항 65는 조항 64의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 프로세서로 하여금, 센서 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하게 한다.

[0191] [0206] 조항 66은 조항 64의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 프로세서로 하여금, 오디오 데이터 샘플들과 연관된 오디오 장면을 식별하는 입력에 기반하여, 사운드 이벤트 분류 모델을 선택하게 한다.

[0192] [0207] 조항 67은 조항 64의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 프로세서로 하여금, 오디오 데이터 샘플들이 수신된 경우에 기반하여 사운드 이벤트 분류 모델을 선택하게 한다.

[0193] [0208] 조항 68은 조항 64의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 프로세서로 하여금, 설정 데이터에 기반하여 사운드 이벤트 분류 모델을 선택하게 한다.

[0194] [0209] 조항 69는 조항 62 내지 68 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하고, 명령들은 프로세서로 하여금, 사운드 클래스가 인식되었다는 결정에 기반하여, 오디오 데이터 샘플들과 연관된 사운드 클

래스를 표시하는 출력을 생성하게 한다.

[0195] [0210] 조항 70은 조항 62 내지 69 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하고, 명령들은 프로세서로 하여금, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하지 않는다는 결정에 기반하여, 오디오 데이터 샘플들에 대응하는 오디오 데이터를 새로운 사운드 이벤트 분류 모델에 대한 트레이닝 데이터로서 저장하게 한다.

[0196] [0211] 조항 71은 조항 62 내지 70 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하고, 명령들은 프로세서로 하여금, 사운드 이벤트 분류 모델은 출력과 연관된 신뢰도 메트릭을 생성하게 하고, 그리고 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부의 결정은 신뢰도 메트릭에 기반한다.

[0197] [0212] 조항 72는 조항 62 내지 71의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 명령들은 프로세서로 하여금, 모델 업데이트 데이터에 기반하여 사운드 이벤트 분류 모델을 업데이트하게 한다.

[0198] [0213] 조항 73은, 조항 62 내지 72 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 오디오 장면을 표시하는 사용자 입력에 기반한다.

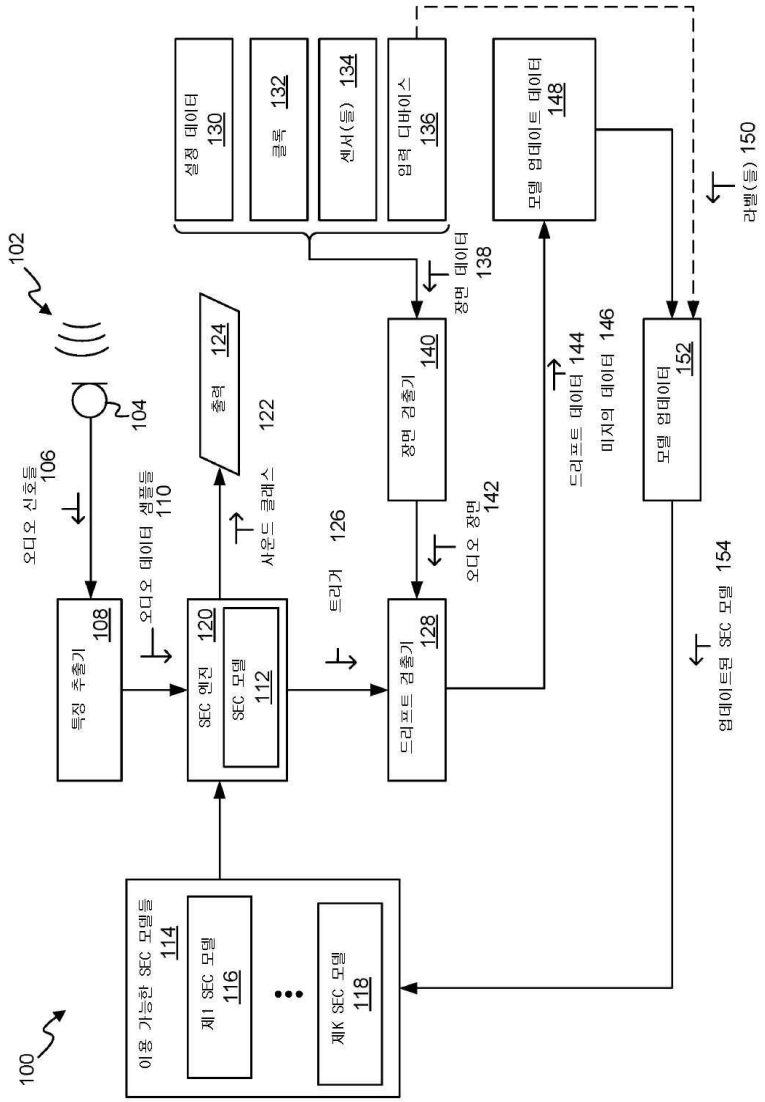
[0199] [0214] 조항 74는 조항 62 내지 73 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 센서 데이터에 기반한다.

[0200] [0215] 조항 75는 조항 62 내지 74 중 어느 한 조항의 비일시적 컴퓨터 판독 가능 매체를 포함하며, 사운드 이벤트 분류 모델이 오디오 장면에 대응하는지 여부의 결정은 오디오 데이터 샘플들과 연관된 타임스탬프에 기반한다.

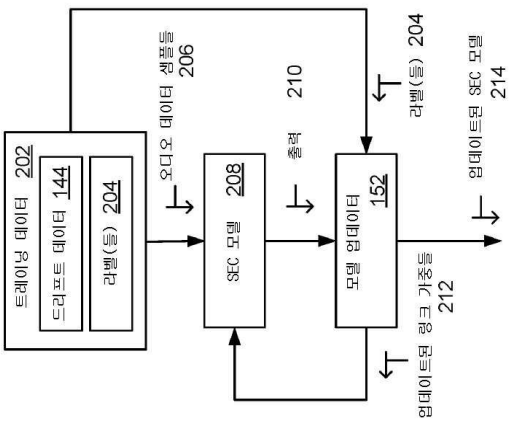
[0201] [0216] 기재된 양상들의 이전 설명은 당업자가 기재된 예들을 사용하거나 또는 실시할 수 있도록 제공된다. 이러한 양상들에 대한 다양한 변형들은 당업자들에게 용이하게 명백할 것이며, 본 명세서에 정의된 원리들은 본 개시의 범위를 벗어나지 않으면서 다른 양상들에 적용될 수 있다. 따라서, 본 개시내용은 본 명세서에 예시된 양상들로 제한되도록 의도되는 것이 아니라, 다음의 청구항들에 의해 정의된 바와 같은 원리들 및 신규한 특성들과 일치하는 가능한 가장 넓은 범위에 부합할 것이다.

## 도면

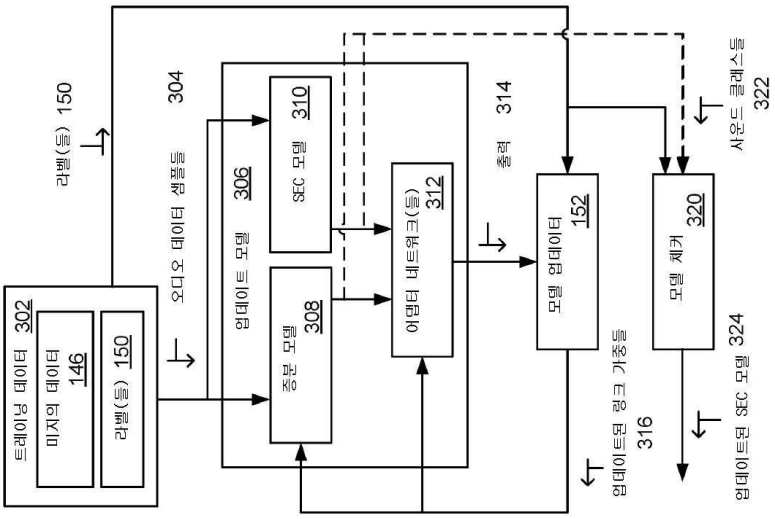
### 도면1


100
102
104
106
108
110
112
114
116
118
120
122
124
126
128
130
132
134
136
138
140
142
144
146
148
150
152
154
이용 가능한 SEC 모델들
제1 SEC 모델
제K SEC 모델
특징 추출기
오디오 데이터 샘플들
SEC 엔진
SEC 모델
사운드 클래스
출력
트리거
드리프트 검출기
오디오 장면
장면 검출기
장면 데이터
설정 데이터
클록
센서(들)
입력 디바이스
드리프트 데이터
미지의 데이터
모델 업데이트 데이터
라벨(들)
모델 업데이터
업데이트된 SEC 모델
오디오 신호들

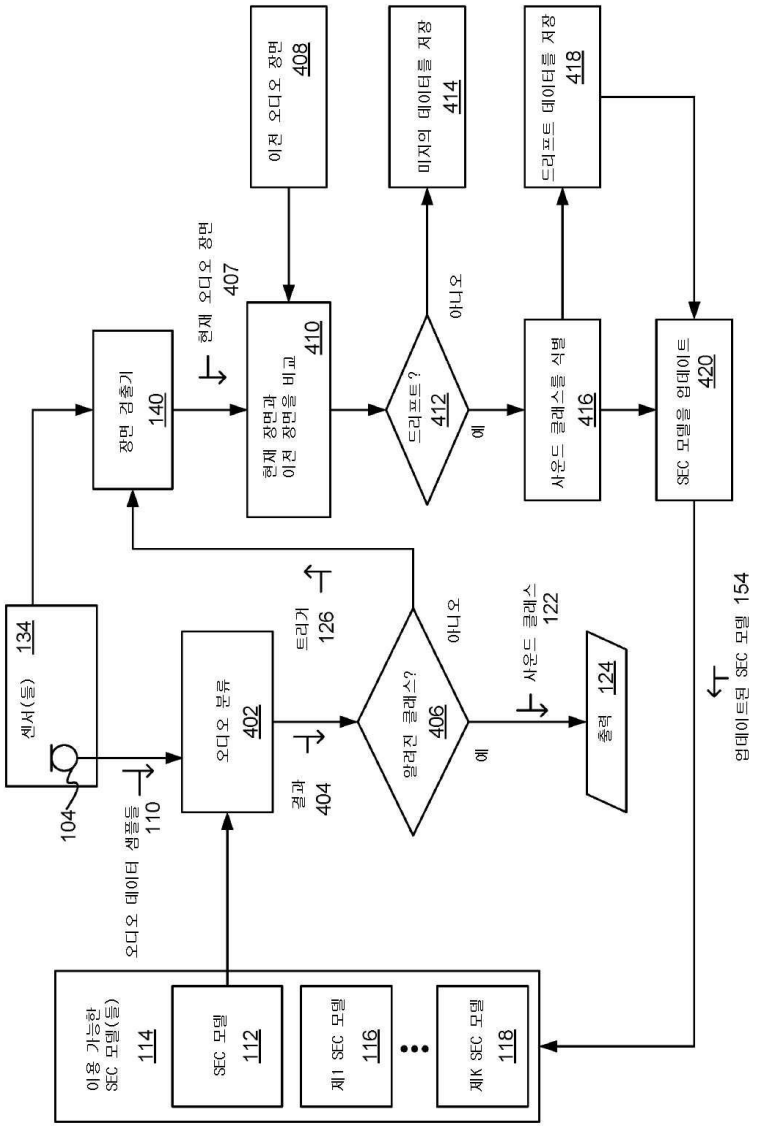
### 도면2


202
204
206
208
210
212
214
트레이닝 데이터
드리프트 데이터
라벨(들)
오디오 데이터 샘플들
SEC 모델
출력
모델 업데이터
업데이트된 링크 가중들
라벨(들)
업데이트된 SEC 모델

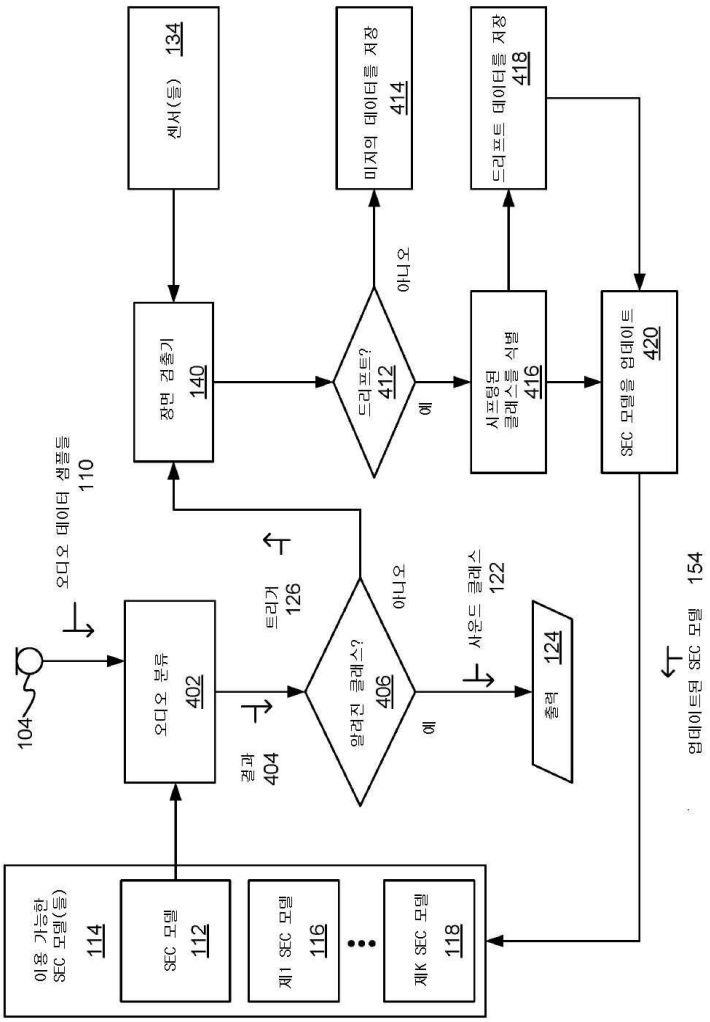
도면3


트레이닝 데이터 302
미지의 데이터 146
라벨(들) 150
라벨(들) 150
오디오 데이터 샘플들 304
업데이트 모델 306
증분 모델 308
SEC 모델 310
어댑터 네트워크(들) 312
출력 314
모델 업데이터 152
업데이트된 링크 가중들 316
모델 체커 320
사운드 클래스들 322
업데이트된 SEC 모델 324

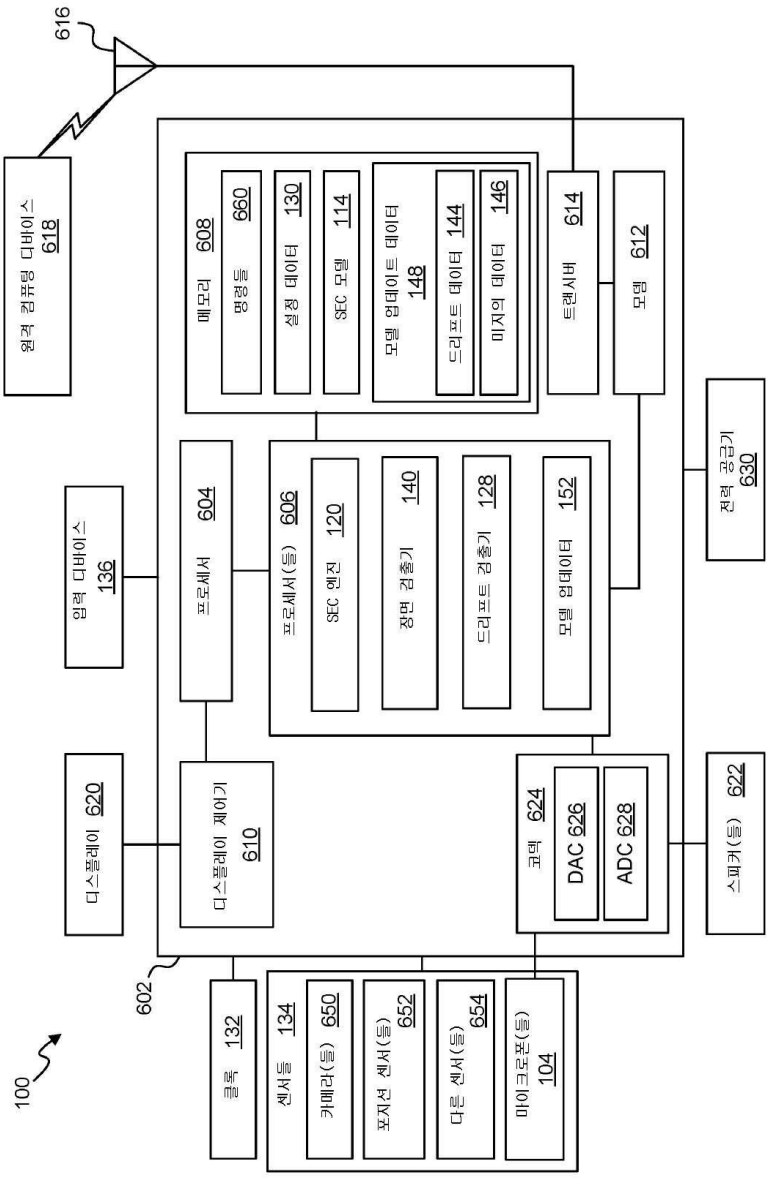
도면4


센서(들) 134
104
오디오 데이터 샘플들 110
오디오 분류 402
결과 404
트리거 126
알려진 클래스? 406
아니오
예
사운드 클래스 122
출력 124
장면 검출기 140
현재 오디오 장면 407
현재 장면과 이전 장면을 비교 410
이전 오디오 장면 408
드리프트? 412
아니오
예
미지의 데이터를 저장 414
사운드 클래스를 식별 416
드리프트 데이터를 저장 418
SEC 모델을 업데이트 420
업데이트된 SEC 모델 154
이용 가능한 SEC 모델(들) 114
SEC 모델 112
제1 SEC 모델 116
제K SEC 모델 118

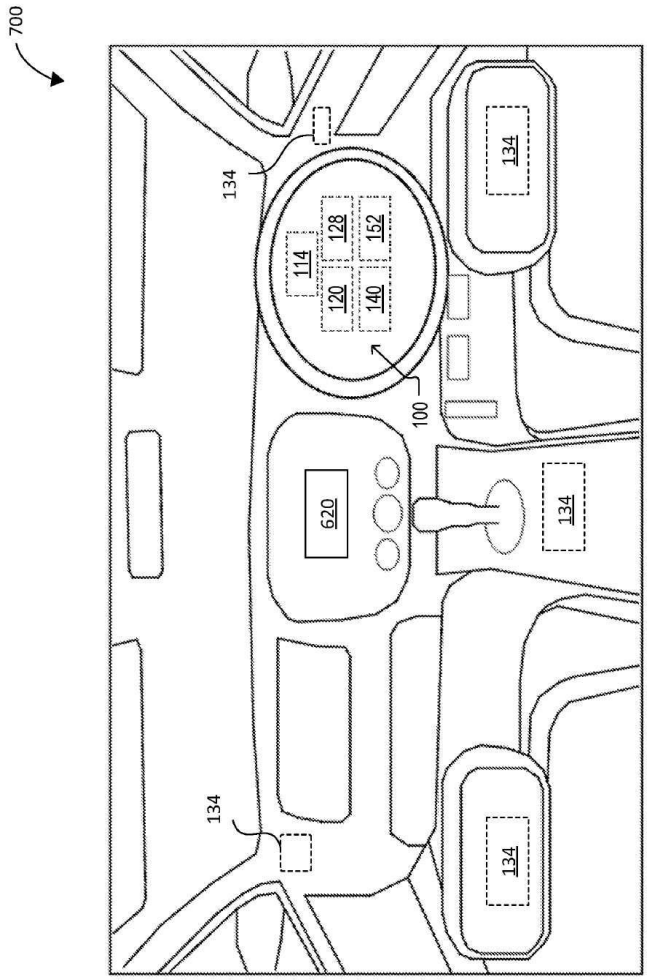
도면5


센서(들) 134
미지의 데이터를 저장 414
드리프트 데이터를 저장 418
정면 검출기 140
드리프트? 412
아니오
예
시프팅된 클래스를 식별 416
SEC 모델을 업데이트 420
오디오 데이터 샘플들 110
104
오디오 분류 402
트리거 126
결과 404
알려진 클래스? 406
아니오
예
사운드 클래스 122
출력 124
업데이트된 SEC 모델 154
이용 가능한 SEC 모델(들) 114
SEC 모델 112
제1 SEC 모델 116
제K SEC 모델 118

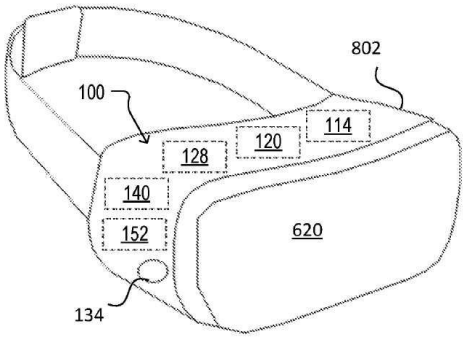
도면6


100
602
616
원격 컴퓨팅 디바이스 618
메모리 608
명령들 660
설정 데이터 130
SEC 모델 114
모델 업데이트 데이터 148
드리프트 데이터 144
미지의 데이터 146
트랜시버 614
모뎀 612
입력 디바이스 136
프로세서 604
프로세서(들) 606
SEC 엔진 120
장면 검출기 140
드리프트 검출기 128
모델 업데이터 152
전력 공급기 630
디스플레이 620
디스플레이 제어기 610
코덱 624
DAC 626
ADC 628
스피커(들) 622
클록 132
센서들 134
카메라(들) 650
포지션 센서(들) 652
다른 센서(들) 654
마이크로폰(들) 104

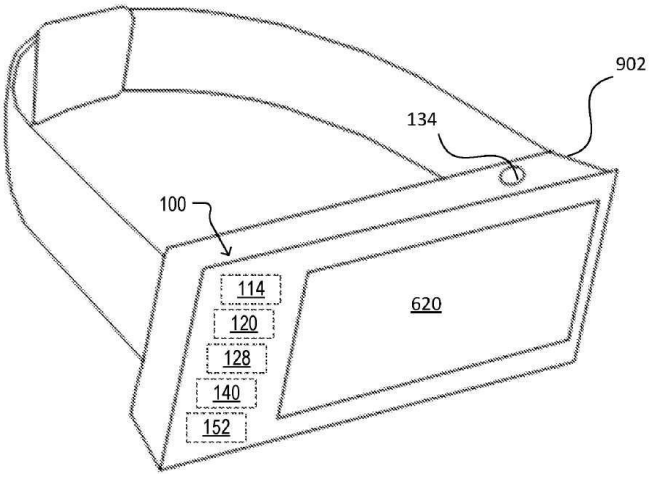
도면7


700
134
114
128
152
120
140
100
620
134
134
134

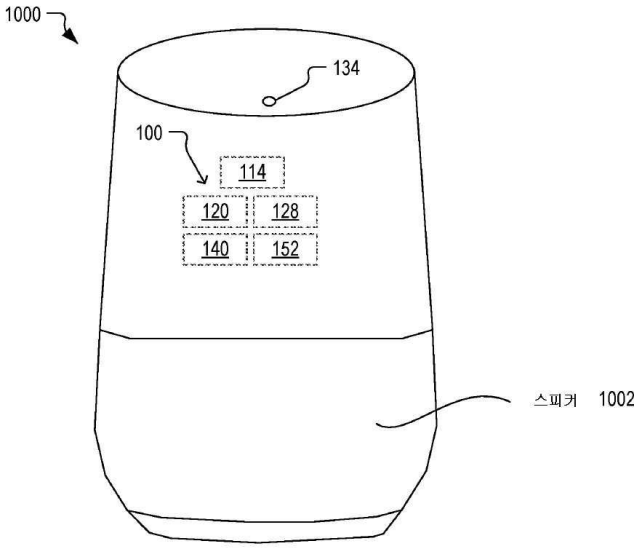
도면8


802
100
114
120
128
140
152
620
134

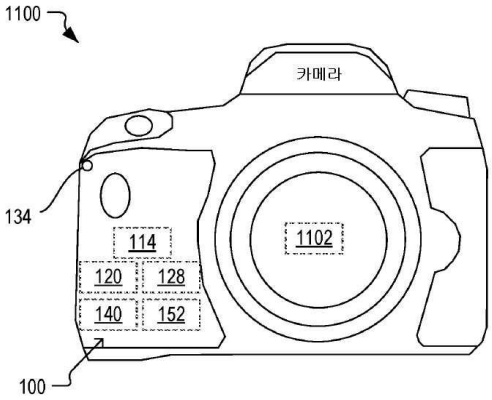
도면9



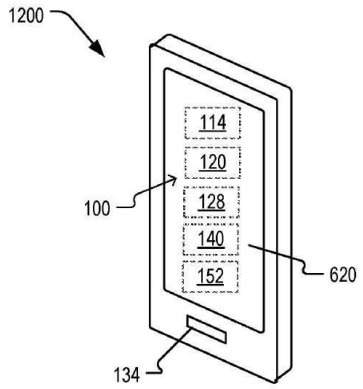

도면10



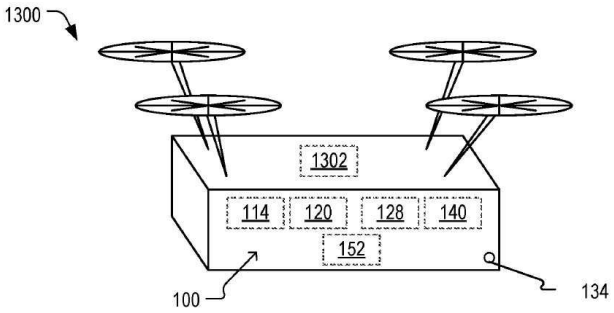

도면11



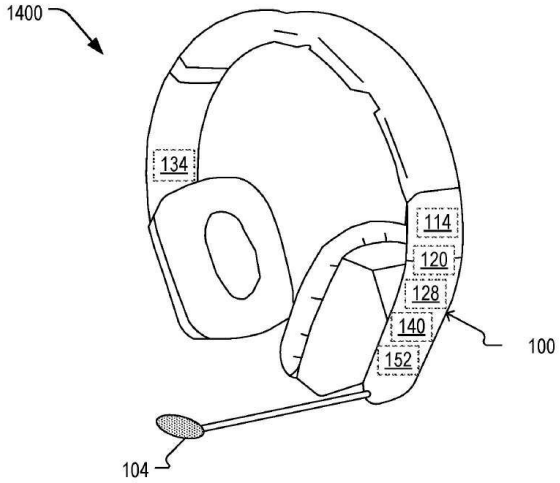
도면12



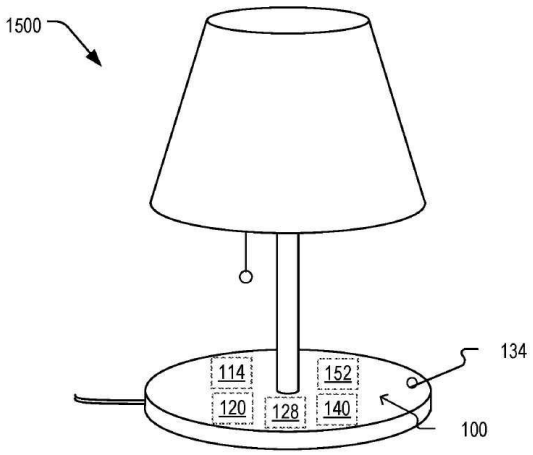

도면13



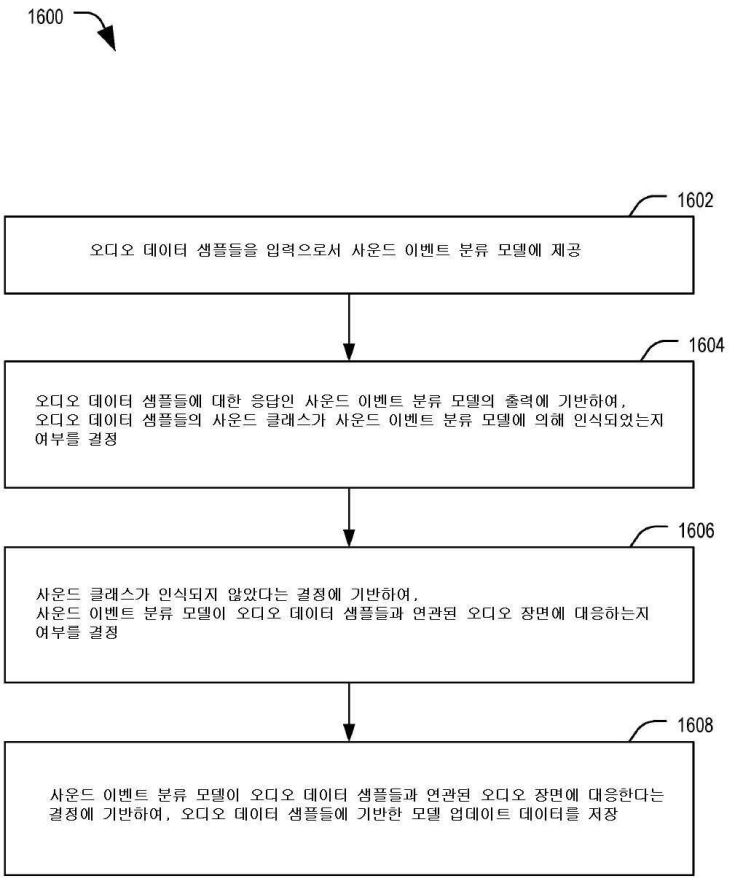

도면14



도면15


1500
134
114
152
120
128
140
100

도면16


1600
1602
오디오 데이터 샘플들을 입력으로서 사운드 이벤트 분류 모델에 제공
1604
오디오 데이터 샘플들에 대한 응답인 사운드 이벤트 분류 모델의 출력에 기반하여, 오디오 데이터 샘플들의 사운드 클래스가 사운드 이벤트 분류 모델에 의해 인식되었는지 여부를 결정
1606
사운드 클래스가 인식되지 않았다는 결정에 기반하여, 사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응하는지 여부를 결정
1608
사운드 이벤트 분류 모델이 오디오 데이터 샘플들과 연관된 오디오 장면에 대응한다는 결정에 기반하여, 오디오 데이터 샘플들에 기반한 모델 업데이트 데이터를 저장