(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0162702
(43) 공개일자 2022년12월08일

(51) 국제특허분류(Int. Cl.)
*G10L 25/78* (2013.01) *G06K 9/62* (2022.01)
*G10L 15/16* (2006.01) *H04W 52/02* (2009.01)
(52) CPC특허분류
*G10L 25/78* (2013.01)
*G06K 9/6228* (2022.01)
(21) 출원번호 10-2022-7032990
(22) 출원일자(국제) 2021년03월01일
심사청구일자 없음
(85) 번역문제출일자 2022년09월22일
(86) 국제출원번호 PCT/US2021/020220
(87) 국제공개번호 WO 2021/202023
국제공개일자 2021년10월07일
(30) 우선권주장
16/837,420 2020년04월01일 미국(US)

(71) 출원인
**퀄컴 인코포레이티드**
미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
(72) 발명자
**쿨카르니, 프라작트**
미국 92121 캘리포니아 샌 디에고 모어하우스 드라이브 5775
**궈, 인이**
미국 92121 캘리포니아 샌 디에고 모어하우스 드라이브 5775
**비서, 에릭**
미국 92121 캘리포니아 샌 디에고 모어하우스 드라이브 5775
(74) 대리인
**특허법인 남앤남**

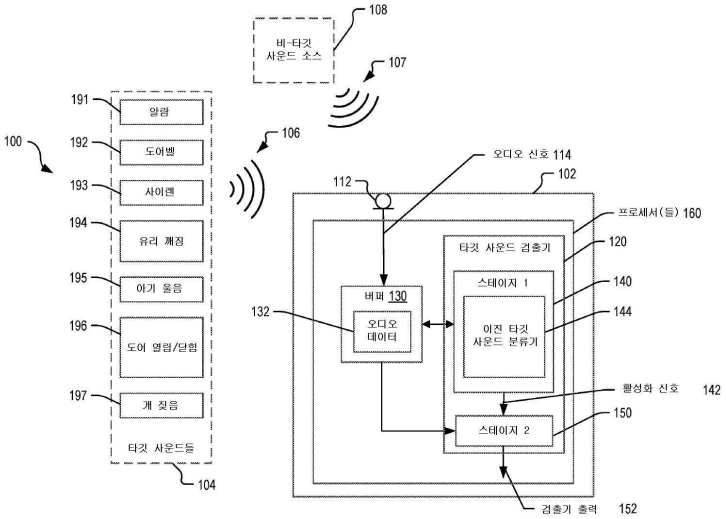
전체 청구항 수 : 총 30 항

(54) 발명의 명칭 **타깃 사운드 검출을 위한 방법 및 장치**

## (57) 요 약

타깃 사운드 검출을 수행하기 위한 디바이스는 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 오디오 데이터를 저장하도록 구성된 버퍼 및 타깃 사운드 검출기를 포함한다. 타깃 사운드 검출기는 제1 스테이지 및 제2 스테이지를 포함한다. 제1 스테이지는 오디오 데이터를 프로세싱하도록 구성된 이진 타깃 사운드 분류기를 포함한다. 제1 스테이지는 타깃 사운드의 검출에 대한 응답으로 제2 스테이지를 활성화하도록 구성된다. 제2 스테이지는 타깃 사운드의 검출에 대한 응답으로 버퍼로부터 오디오 데이터를 수신하도록 구성된다.

*대 표 도* - 도1



(52) CPC특허분류
*G06K 9/6268* (2022.01)
*G10L 15/16* (2013.01)
*H04W 52/0229* (2013.01)
*H04W 52/0261* (2013.01)

# 명 세 서

## 청구범위

### 청구항 1

사운드 검출을 수행하기 위한 디바이스로서,

하나 이상의 프로세서들을 포함하고, 상기 하나 이상의 프로세서들은,

오디오 데이터를 저장하도록 구성된 버퍼; 및

제1 스테이지 및 제2 스테이지를 포함하는 타깃 사운드 검출기를 포함하며,

상기 제1 스테이지는 상기 오디오 데이터를 프로세싱하도록 구성된 이진 타깃 사운드 분류기를 포함하고, 상기 제1 스테이지는 상기 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 상기 제2 스테이지를 활성화하도록 구성되고, 그리고 상기 제2 스테이지는 상기 타깃 사운드의 검출에 대한 응답으로 상기 버퍼로부터 상기 오디오 데이터를 수신하도록 구성되는,

사운드 검출을 수행하기 위한 디바이스.

### 청구항 2

제1 항에 있어서,

상기 이진 타깃 사운드 분류기는 제1 값 및 제2 값을 포함하는 이진 신호를 생성하도록 추가로 구성되고,

상기 제1 값은 상기 오디오 데이터에서 다수의 타깃 사운드들 중 임의의 타깃 사운드의 존재를 검출하는 것에 대한 응답으로 상기 제2 스테이지를 활성화하도록 설정되고; 그리고

상기 제2 값은 상기 다수의 타깃 사운드들 중 어느 것도 상기 오디오 데이터에 있지 않음을 검출하는 것에 대한 응답으로 상기 신호를 생성하는 것을 억제하도록 설정되는,

사운드 검출을 수행하기 위한 디바이스.

### 청구항 3

제1 항에 있어서,

상기 이진 타깃 사운드 분류기는 뉴럴 네트워크를 포함하는,

사운드 검출을 수행하기 위한 디바이스.

### 청구항 4

제1 항에 있어서,

상기 이진 타깃 사운드 분류기는 베이지안(Bayesian) 분류기 또는 GMM(Gaussian Mixed Model) 분류기 중 적어도 하나를 포함하는,

사운드 검출을 수행하기 위한 디바이스.

### 청구항 5

제1 항에 있어서,

상기 제2 스테이지는, 다수의 타깃 사운드들 각각에 대해, 상기 오디오 데이터에서 그 타깃 사운드의 존재 또는 부재를 표시하는 검출기 출력을 생성하도록 구성된 다중 타깃 사운드 분류기를 포함하고, 상기 다수의 타깃 사운드들은 사운드 이벤트들의 다수의 클래스들에 대응하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 6**

제1 항에 있어서,

상기 이진 타깃 사운드 분류기 및 상기 버퍼는 저전력 도메인에 포함되고 올웨이즈-온(always-on) 모드로 동작하도록 구성되며, 상기 제2 스테이지는 상기 신호를 수신하는 것에 응답하여 저전력 상태로부터 활성 상태로 전환하도록 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 7**

제1 항에 있어서,

상기 신호는 웨이크업 인터럽트 신호에 대응하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 8**

제1 항에 있어서,

상기 제1 스테이지는 상기 제1 스테이지에 의한 상기 타깃 사운드의 검출에 대한 응답으로 카메라를 활성화하도록 추가로 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 9**

제8 항에 있어서,

상기 카메라로부터의 입력 신호에 적어도 부분적으로 기반하여 상기 디바이스의 환경을 분류하도록 구성된 장면 검출기를 더 포함하고,

상기 제2 스테이지는 사운드 이벤트들의 다수의 클래스들 가운데서 상기 오디오 데이터를 분류하도록 구성되는 다중 타깃 사운드 분류기를 포함하고, 상기 다중 타깃 사운드 분류기의 동작은 상기 장면 검출기에 의해 분류되는 상기 환경에 적어도 부분적으로 기반하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 10**

제9 항에 있어서,

상기 다중 타깃 사운드 분류기는 상기 환경에 대응하는 상기 사운드 이벤트들의 다수의 클래스들 중 하나 이상의 특정 클래스들에 포커싱하도록(focus on) 조정되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 11**

제9 항에 있어서,

상기 다중 타깃 사운드 분류기는,

사운드 이벤트 클래스들의 다수의 세트들 가운데서 상기 환경에 대응하는 사운드 이벤트 클래스들의 특정 세트를 선택하고; 그리고

상기 특정 세트의 상기 사운드 이벤트 클래스들에 기반하여 상기 오디오 데이터를 분류하도록 추가로 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 12**

제1 항에 있어서,

상기 타깃 사운드 검출기는, 트레이닝된 데이터의 하나 이상의 세트들 가운데서 상기 디바이스의 검출된 환경에 대응하는 트레이닝된 데이터의 특정 세트를 선택하고, 그리고 상기 트레이닝된 데이터의 특정 세트에 기반하여 상기 오디오 데이터를 프로세싱하도록 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 13**

제12 항에 있어서,

상기 환경은 카메라, 위치 검출 시스템, 또는 오디오 장면 검출기 중 적어도 하나에 기반하여 검출되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 14**

제1 항에 있어서,

상기 이진 타깃 사운드 분류기에 의해 상기 오디오 데이터에서 다수의 타깃 사운드들 중 임의의 사운드의 존재를 검출하는 것에 응답하여 활성화되도록 구성된 오디오 장면 검출기를 더 포함하고,

상기 오디오 장면 검출기는,

상기 오디오 데이터를 프로세싱하고 오디오 장면 변화의 검출에 대한 응답으로 장면 변화 신호를 생성하도록 구성된 오디오 장면 변화 검출기; 및

상기 오디오 장면 변화의 검출에 대한 응답으로 상기 버퍼로부터 상기 오디오 데이터를 수신하도록 구성된 오디오 장면 분류기를 포함하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 15**

제14 항에 있어서,

상기 오디오 장면 분류기는 다수의 오디오 장면 클래스들에 따라 상기 오디오 데이터를 분류하도록 구성되고, 상기 다수의 오디오 장면 클래스들은, 집에서, 사무실에서, 식당에서, 자동차에서, 기차에서, 거리에서, 실내에서, 또는 실외에서 중 적어도 2개를 포함하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 16**

제14 항에 있어서,

상기 오디오 장면 변화 검출기는 소음 통계(noise statistics) 또는 비-정지형(non-stationary) 사운드 통계 중 적어도 하나에서의 변화들을 검출하는 것에 기반하여 상기 오디오 장면 변화를 검출하도록 추가로 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 17**

제14 항에 있어서,

상기 오디오 장면 변화 검출기는 장면들 사이의 전환들에 대응하는 오디오 데이터를 사용하여 트레이닝된 분류기를 포함하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 18**

제14 항에 있어서,

상기 오디오 장면 검출기는 계층적 검출기에 대응하고;

상기 오디오 장면 변화 검출기는 오디오 장면 클래스들의 제1 세트 내의 오디오 장면 클래스들 사이의 변화들을 검출하는 것에 기반하여 상기 오디오 장면 변화를 검출하도록 구성되고; 그리고

상기 오디오 장면 분류기는 오디오 장면 클래스들의 제2 세트에 따라 상기 오디오 데이터를 분류하도록 구성되며,

상기 오디오 장면 클래스들의 제1 세트 내의 상기 오디오 장면 클래스들의 제1 카운트는 상기 오디오 장면 클래스들의 제2 세트 내의 상기 오디오 장면 클래스들의 제2 카운트보다 적은,

사운드 검출을 수행하기 위한 디바이스.

**청구항 19**

제1 항에 있어서,

상기 하나 이상의 프로세서들에 커플링되고 상기 오디오 데이터를 생성하도록 구성된 마이크로폰을 더 포함하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 20**

제1 항에 있어서,

상기 제2 스테이지는, 다수의 타깃 사운드들 각각에 대해, 상기 오디오 데이터에서 그 타깃 사운드의 존재 또는 부재를 표시하는 검출기 출력을 생성하도록 구성된 다중 타깃 사운드 분류기를 포함하고, 상기 다수의 타깃 사운드들은, 운송수단 도어 열림 또는 닫힘, 도로 소음, 창문 열림 또는 닫힘, 제동, 핸드 브레이크 맞물림 또는 맞물림해제, 앞유리 와이퍼(windshield wiper)들, 방향 지시등, 또는 엔진 회전 중 하나 이상에 대응하는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 21**

제1 항에 있어서,

상기 하나 이상의 프로세서들은 무선 스피커 및 음성 활성화 디바이스(voice activated device) 내에 구현되고, 상기 무선 스피커 및 음성 활성화 디바이스는,

통합형 어시스턴트 애플리케이션; 및

상기 통합형 어시스턴트 애플리케이션에 응답하여 활성화되도록 구성된 카메라를 포함하고, 상기 카메라는 상기 이진 타깃 사운드 분류기에 의해 상기 오디오 데이터에서 다수의 타깃 사운드들 중 임의의 타깃 사운드의 존재의 검출에 응답하여 활성화되도록 추가로 구성되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 22**

제1 항에 있어서,

상기 하나 이상의 프로세서들은 휴대가능 전자 디바이스 내에 구현되는,

사운드 검출을 수행하기 위한 디바이스.

**청구항 23**

타깃 사운드 검출 방법으로서,

버퍼에 오디오 데이터를 저장하는 단계:

타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여 상기 버퍼 내의 상기 오디오 데이터를 프로세싱하는 단계;

상기 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 상기 타깃 사운드 검출기의 제2 스테이지를 활

성화하는 단계; 및

상기 제2 스테이지에서 다중 타깃 사운드 분류기를 사용하여 상기 버퍼로부터의 상기 오디오 데이터를 프로세싱하는 단계를 포함하는,

타깃 사운드 검출 방법.

**청구항 24**

제23 항에 있어서,

상기 이진 타깃 사운드 분류기 및 상기 버퍼는 올웨이즈-온 모드로 동작하고,

상기 제2 스테이지를 활성화하는 단계는, 상기 제1 스테이지로부터 상기 제2 스테이지로 신호를 전송하는 단계 및 상기 제2 스테이지에서 상기 신호를 수신하는 것에 응답하여 상기 제2 스테이지를 저전력 상태로부터 활성 상태로 전환하는 단계를 포함하는,

타깃 사운드 검출 방법.

**청구항 25**

제23 항에 있어서,

상기 다중 타깃 사운드 분류기는 사운드 이벤트들의 다수의 클래스들에 대응하는 다수의 타깃 사운드들에 기반하여 상기 오디오 데이터를 프로세싱하는,

타깃 사운드 검출 방법.

**청구항 26**

제23 항에 있어서,

장면 분류기에서, 다수의 오디오 장면 클래스들에 따라 상기 오디오 데이터를 분류하는 단계를 더 포함하고, 상기 다수의 오디오 장면 클래스들은, 집에서, 사무실에서, 식당에서, 자동차에서, 기차에서, 거리에서, 실내에서, 또는 실외에서 중 적어도 2개를 포함하는,

타깃 사운드 검출 방법.

**청구항 27**

제23 항에 있어서,

오디오 장면 클래스들의 제1 세트 내의 오디오 장면 클래스들 사이의 변화들을 검출하는 것에 기반하여 오디오 장면 변화를 검출하기 위해 상기 오디오 데이터를 프로세싱하는 단계; 및

오디오 장면 클래스들의 제2 세트에 기반하여 상기 오디오 데이터를 분류하는 단계를 더 포함하고,

상기 오디오 장면 클래스들의 제1 세트 내의 상기 오디오 장면 클래스들의 제1 카운트는 상기 오디오 장면 클래스들의 제2 세트 내의 오디오 장면 클래스들의 제2 카운트보다 더 적은,

타깃 사운드 검출 방법.

**청구항 28**

명령들을 저장하는 컴퓨터-판독가능 저장 디바이스로서,

상기 명령들은 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금,

버퍼에 오디오 데이터를 저장하게 하고;

타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여 상기 버퍼 내의 상기 오디오 데이터를 프로세싱하게 하고;

상기 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 상기 타깃 사운드 검출기의 제2 스테이지를 활성화하게 하고;

상기 제2 스테이지에서 다중 타깃 사운드 분류기를 사용하여 상기 버퍼로부터의 상기 오디오 데이터를 프로세싱하게 하는,

컴퓨터-판독가능 저장 디바이스.

**청구항 29**

장치로서,

타깃 사운드를 검출하기 위한 수단 - 상기 타깃 사운드를 검출하기 위한 수단은 제1 스테이지 및 제2 스테이지를 포함하고, 상기 제1 스테이지는, 오디오 데이터의 이진 타깃 사운드 분류를 생성하고, 그리고 상기 오디오 데이터를 상기 타깃 사운드를 포함하는 것으로서 분류하는 것에 대한 응답으로 상기 제2 스테이지를 활성화하기 위한 수단을 포함함 -; 및

상기 오디오 데이터를 버퍼링하고, 그리고 타깃 사운드를 포함하는 것으로서의 상기 오디오 데이터의 분류에 대한 응답으로 상기 오디오 데이터를 상기 제2 스테이지에 제공하기 위한 수단을 포함하는,

장치.

**청구항 30**

제29 항에 있어서,

오디오 장면을 검출하기 위한 수단을 더 포함하고, 상기 오디오 장면을 검출하기 위한 수단은,

상기 오디오 데이터에서 오디오 장면 변화를 검출하기 위한 수단; 및

상기 오디오 장면 변화의 검출에 대한 응답으로 상기 오디오 데이터를 특정 오디오 장면으로서 분류하기 위한 수단을 포함하는,

장치.

## *발명의 설명*

### *기 술 분 야*

[0001] [0001] 본 출원은 2020년 4월 1일자로 출원된 공동 소유의 미국 정규 특허 출원 제16/837,420호를 우선권으로 주장하며, 상기 출원은 그 전체가 인용에 의해 통합된다.

[0002] [0002] 본 개시내용은 일반적으로 오디오 데이터에서의 타깃 사운드들의 검출에 관한 것이다.

### *배 경 기 술*

[0003] [0003] 오디오 콘텍스트 검출은 전자 디바이스가 전자 디바이스에 의해 캡처된 오디오에 기반하여 맥락 정보(contextual information)를 식별할 수 있게 하는 데 통상적으로 사용된다. 예컨대, 전자 디바이스는 수신된 사운드를 분석하여 사운드가 미리 결정된 사운드 이벤트를 표시하는지 여부를 결정할 수 있다. 다른 예로서, 전자 디바이스는 주변 환경, 이를테면 집 환경 또는 사무실 환경을 분류하기 위해 수신된 사운드를 분석할 수 있다. "올웨이즈-온(always-on)" 오디오 콘텍스트 검출 시스템은 전자 디바이스가 오디오 입력을 지속적으로 스캔하여 오디오 입력에서 사운드 이벤트들을 검출할 수 있게 한다. 그러나, 오디오 콘텍스트 검출 시스템의 연속적인 동작은 비교적 큰 전력 소비를 초래하며, 이는 모바일 디바이스에서 구현되는 경우 배터리 수명을 감소시킨다. 또한, 오디오 콘텍스트 검출 시스템이 검출하도록 구성된 사운드 이벤트들의 수가 증가함에 따라 시스템 복잡성 및 전력 소비가 증가한다.

### *발명의 내용*

[0004] [0004] 본 개시내용의 일 구현에 따르면, 사운드 검출을 수행하기 위한 디바이스는 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은 오디오 데이터를 저장하도록 구성된 버퍼를 포함한다. 하나 이상의 버퍼들은 또한, 제1 스테이지 및 제2 스테이지를 포함하는 타깃 사운드 검출기를 포함한다. 제1 스테이지는 오디오 데이터를 프로세싱하도록 구성된 이진 타깃 사운드 분류기를 포함한다. 제1 스테이지는 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 제2 스테이지를 활성화하도록 구성된다. 제2 스테이지는 타깃 사운드의

검출에 대한 응답으로 버퍼로부터 오디오 데이터를 수신하도록 구성된다.

[0005] [0005] 본 개시내용의 다른 구현에 따르면, 타깃 사운드 검출 방법은 오디오 데이터를 버퍼에 저장하는 단계를 포함한다. 방법은 또한, 타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여 버퍼 내의 오디오 데이터를 프로세싱하는 단계, 및 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 타깃 사운드 검출기의 제2 스테이지를 활성화하는 단계를 포함한다. 방법은 제2 스테이지에서 다중 타깃 사운드 분류기를 사용하여 버퍼로부터의 오디오 데이터를 프로세싱하는 단계를 더 포함한다.

[0006] [0006] 본 개시내용의 다른 구현에 따르면, 컴퓨터-판독가능 저장 디바이스는 명령들을 저장하며, 명령들은 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 오디오 데이터를 버퍼에 저장하게 하고, 그리고 타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여 버퍼 내의 오디오 데이터를 프로세싱하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때 또한, 하나 이상의 프로세서들로 하여금, 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 타깃 사운드 검출기의 제2 스테이지를 활성화하게 하고, 그리고 제2 스테이지에서 다중 타깃 사운드 분류기를 사용하여 버퍼로부터의 오디오 데이터를 프로세싱하게 한다.

[0007] [0007] 본 개시내용의 다른 구현에 따르면, 장치는 타깃 사운드를 검출하기 위한 수단을 포함한다. 타깃 사운드를 검출하기 위한 수단은 제1 스테이지 및 제2 스테이지를 포함한다. 제1 스테이지는, 오디오 데이터의 이진 타깃 사운드 분류를 생성하고, 그리고 타깃 사운드를 포함하는 것으로서 오디오 데이터를 분류하는 것에 대한 응답으로 제2 스테이지를 활성화하기 위한 수단을 포함한다. 장치는 또한, 오디오 데이터를 버퍼링하고, 그리고 타깃 사운드를 포함하는 것으로서의 오디오 데이터의 분류에 대한 응답으로 오디오 데이터를 제2 스테이지에 제공하기 위한 수단을 포함한다.

## *도면의 간단한 설명*

[0008] [0008] 도 1은 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 디바이스를 포함하는 시스템의 특정한 예시적인 구현의 도면이다.

[0009] [0009] 도 2는 본 개시내용의 일부 예들에 따른, 도 1의 디바이스의 특정한 구현의 다이어그램이다.

[0010] [0010] 도 3은 본 개시내용의 일부 예들에 따른, 멀티-스테이지 오디오 장면 검출기를 포함하는 도 1의 디바이스의 다른 특정한 구현의 다이어그램이다.

[0011] [0011] 도 4는 본 개시내용의 일부 예들에 따른, 멀티-스테이지 오디오 장면 검출기에 통합될 수 있는 컴포넌트들의 특정 예의 다이어그램이다.

[0012] [0012] 도 5는 본 개시내용의 일부 예들에 따른, 멀티-스테이지 오디오 장면 검출기에 통합될 수 있는 컴포넌트들의 다른 특정한 예의 다이어그램이다.

[0013] [0013] 도 6은 본 개시내용의 일부 예들에 따른, 장면 검출기를 포함하는 도 1의 디바이스의 다른 특정한 구현의 다이어그램이다.

[0014] [0014] 도 7은 본 개시내용의 일부 예들에 따른, 도 6의 디바이스에 통합될 수 있는 컴포넌트들의 특정한 예의 다이어그램이다.

[0015] [0015] 도 8은 본 개시내용의 일부 예들에 따른, 도 6의 디바이스에 통합될 수 있는 컴포넌트들의 다른 특정한 예의 다이어그램이다.

[0016] [0016] 도 9는 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 집적 회로의 예를 예시한다.

[0017] [0017] 도 10은 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 운송수단의 제1 예의 다이어그램이다.

[0018] [0018] 도 11은 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 운송수단의 제2 예의 다이어그램이다.

[0019] [0019] 도 12는 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 헤드셋, 이를테면 가상 현실 또는 증강 현실 헤드셋의 다이어그램이다.

[0020] 도 13은 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 웨어러블 전자 디바이스의 다이어그램이다.

[0021] 도 14는 본 개시내용의 일부 예들에 따른, 멀티-스테이지 타깃 사운드 검출기를 포함하는 음성-제어 스피커 시스템의 다이어그램이다.

[0022] 도 15는 본 개시내용의 일부 예들에 따른, 도 1의 디바이스에 의해 수행될 수 있는 타깃 사운드 검출의 방법의 특정한 구현의 다이어그램이다.

[0023] 도 16은 본 개시내용의 일부 예들에 따른, 타깃 사운드 검출을 수행하도록 동작가능한 디바이스의 특정한 예시적인 예의 블록 다이어그램이다.

## *발명을 실시하기 위한 구체적인 내용*

[0009] [0024] 전력 소비를 감소시키기 위해 멀티-스테이지 타깃 사운드 검출기를 사용하는 디바이스들 및 방법들이 개시된다. 오디오 입력에서 오디오 이벤트들을 검출하기 위해 오디오 입력을 지속적으로 스캔하는 올웨이즈-온 사운드 검출 시스템은 비교적 큰 전력 소비를 초래하기 때문에, 올웨이즈-온 사운드 검출 시스템이 전력-제약된 환경, 이를테면 모바일 디바이스에서 구현될 때, 배터리 수명은 감소된다. 사운드 검출 시스템이 검출하도록 구성된 오디오 이벤트들의 수를 감소시킴으로써 전력 소비가 감소될 수 있지만, 오디오 이벤트들의 수를 감소시키는 것은 사운드 검출 시스템의 효용을 감소시킨다.

[0010] [0025] 본원에서 설명된 바와 같이, 멀티-스테이지 타깃 사운드 검출기는 올웨이즈-온 동작을 위해 비교적 저전력을 사용하여 비교적 많은 수의 관심 타깃 사운드들의 검출을 지원한다. 멀티-스테이지 타깃 사운드 검출기는, (그룹으로서) 모든 관심 타깃 사운드들과 비-타깃 사운드들 사이의 오디오 데이터의 이진 분류를 지원하는 제1 스테이지를 포함한다. 멀티-스테이지 타깃 사운드 검출기는, 추가 분석을 수행하고 오디오 데이터를 특정한 하나 이상의 관심 타깃 사운드들을 포함하는 것으로서 범주화하기 위한 제2 스테이지를 포함한다. 제1 스테이지의 이진 분류는 올웨이즈-온 동작 상태에서 사운드 이벤트 검출을 지원하기 위한 낮은 복잡도 및 작은 메모리 풋프린트로 인해 저전력 소비를 가능하게 한다. 제2 스테이지는, 타깃 사운드들 간의 구별을 하고 제1 스테이지에 의해 생성될 수 있는 허위 포지티브들(false positives)(예컨대, 타깃 사운드의 부정확한 검출들)을 감소 또는 제거하기 위한 더 강력한 타깃 사운드 분류기를 포함한다.

[0011] [0026] 일부 구현들에서, 오디오 데이터에서 하나 이상의 관심 타깃 사운드들을 검출하는 것에 대한 응답으로, 제2 스테이지는 오디오 데이터의 더 강력한 프로세싱을 가능하게 하기 위해 (예컨대, 슬립 상태로부터) 활성화된다. 제2 스테이지에서 오디오 데이터의 프로세싱의 완료 시에, 제2 스테이지는 저전력 상태로 리턴할 수 있다. 올웨이즈-온 동작을 위해 제1 스테이지의 낮은 복잡도 이진 분류를 사용하고 제2 스테이지의 더 강력한 타깃 사운드 분류기를 선택적으로 활성화함으로써, 타깃 사운드 검출기는 올웨이즈-온 동작에 대해 감소된 평균 전력 소비를 갖는 고-성능 타깃 사운드 분류를 가능하게 한다.

[0012] [0027] 일부 구현들에서, 다중-스테이지 환경 장면 검출기는 환경 장면 변화가 발생했는지 여부를 검출하는 올웨이즈-온 제1 스테이지를 포함하고, 제1 스테이지가 환경에서의 변화를 검출할 때 선택적으로 활성화되는 더 강력한 제2 스테이지를 또한 포함한다. 일부 예들에서, 제1 스테이지는 임의의 특정 환경 장면을 식별하지 않으면서 오디오 데이터가 환경 장면 변화를 나타내는지 여부를 검출하도록 구성된 이진 분류기를 포함한다. 다른 예들에서, 계층적 장면 변화 검출기는 제1 스테이지에서 비교적 적은 수의 넓은 클래스들(예컨대, 실내에서, 실외에서, 및 운송수단에서)을 검출하도록 구성된 분류기를 포함하고, 제2 스테이지의 더 강력한 분류기는 더 많은 수의 보다 특정한 환경 장면들(예컨대, 자동차에서, 기차에서, 집에서, 사무실에서 등)을 검출하도록 구성된다. 결과적으로, 멀티-스테이지 타깃 사운드 검출에서와 유사한 방식으로 올웨이즈-온 동작에 대해 감소된 평균 전력 소비로 고성능 환경 장면 검출이 제공될 수 있다.

[0013] [0028] 일부 구현들에서, 타깃 사운드 검출기는 자신의 환경에 기반하여 동작을 조정한다. 예컨대, 타깃 사운드 검출기가 사용자의 집 안에 있을 때, 타깃 사운드 검출기는 개 짖음 또는 도어벨과 같은 가정 사운드들과 연관된 트레이닝된 데이터를 사용할 수 있다. 타깃 사운드 검출기가 운송수단, 이를테면 자동차 내에 있을 때, 타깃 사운드 검출기는 유리 깨짐 또는 사이렌과 같은 운송수단 사운드들과 연관된 트레이닝된 데이터를 사용할 수 있다. 이를테면, 오디오 장면 검출기, 카메라, (예컨대, 위성-기반 포지셔닝 시스템으로부터의) 위치 데이터, 또는 기법들의 조합들을 사용하는 다양한 기법들이 환경을 결정하기 위해 사용될 수 있다. 일부 예들에서, 타깃 사운드 검출기의 제1 스테이지는 환경을 결정하기 위해 카메라 또는 다른 컴포넌트를 활성화하고, 타깃 사운드 검출기의 제2 스테이지는 검출된 환경과 연관된 타깃 사운드들의 보다 정확한 검출을 위해 "튜닝"된다.

환경 검출을 위해 카메라 또는 다른 컴포넌트를 사용하는 것은 향상된 타깃 사운드 검출을 가능하게 하고, 타깃 사운드 검출기의 제1 스테이지에 의해 활성화될 때까지 카메라 또는 다른 컴포넌트를 저전력 상태로 유지하는 것은 감소된 전력 소비를 가능하게 한다.

[0014] [0029] “프로듀싱하는"이라는 용어는, 그 맥락에 의해 명시적으로 제한되지 않는 한, 계산하는, 생성하는, 및/또는 제공하는 것과 같은 그것의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다. "제공하는"이라는 용어는, 그 맥락에 의해 명시적으로 제한되지 않는 한, 계산하는, 생성하는, 및/또는 프로듀싱하는 것과 같은 그것의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다. “커플링된"이라는 용어는, 그 맥락에 의해 명시적으로 제한되지 않는 한, 직접적인 또는 간접적인 전기적 또는 물리적 연결을 나타내기 위해 사용된다. 연결이 간접적인 경우, 구조들 사이에 다른 블록들 또는 컴포넌트들이 "커플링"될 수 있다. 예컨대, 라우드스피커는, 라우드스피커로부터 벽으로의 파(wave)들(예컨대, 사운드)의 전파를 가능하게 하는 개재 매체(예컨대, 공기)를 통해 인근의 벽에 음향적으로 커플링될 수 있다(또는 그 반대로도 가능함).

[0015] [0030] 용어 "구성"은 그 특정 맥락에 의해 표시된 바와 같이 방법, 장치, 디바이스, 시스템 또는 이들의 임의의 조합을 참조하여 사용될 수 있다. 본 설명 및 청구항들에서 "포함하는"이라는 용어가 사용되는 경우, 이는 다른 엘리먼트들 또는 동작들을 배제하지 않는다. ("A는 B에 기반한다"의 경우에서와 같은) "~에 기반하는"이라는 용어는, (i) "적어도 ~에 기반하는"(예컨대, "A는 적어도 B에 기반한다") 및, 특정 맥락에서 적절한 경우, (ii) "~와 동일한"(예컨대, "A는 B와 동일하다")의 경우들을 포함하는, 그의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다. (i) A가 B에 기반하는 것이 적어도 기반하는 것을 포함하는 경우에, 이는 A가 B에 커플링되는 구성을 포함할 수 있다. 유사하게, "~에 응답하여"라는 용어는 "적어도 ~에 응답하여"를 포함하는 그의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다. "적어도 하나"라는 용어는 "하나 이상"을 포함하는 그의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다. "적어도 2개"라는 용어는 "2개 이상"을 포함하는 그의 통상적인 의미들 중 임의의 것을 나타내기 위해 사용된다.

[0016] [0031] "장치" 및 "디바이스"라는 용어들은, 특정 맥락에 의해 달리 표시되지 않는 한 일반적으로 그리고 상호교환가능하게 사용된다. 달리 표시되지 않는 한, 특정 피처를 갖는 장치의 동작에 대한 임의의 개시내용은 또한, 유사한 피처를 갖는 방법을 개시하도록 명시적으로 의도되며(그리고 그 반대도 가능함), 특정 구성에 따른 장치의 동작에 대한 임의의 개시내용은 또한, 유사한 구성에 따른 방법을 개시하도록 명시적으로 의도된다(그리고 그 반대도 가능함). "방법", "프로세스", "절차" 및 "기법"이라는 용어들은, 특정 맥락에서 달리 표시되지 않는 한 일반적으로 그리고 상호교환가능하게 사용된다. "엘리먼트" 및 "모듈"이라는 용어들은 더 큰 구성의 일부를 나타내기 위해 사용될 수 있다. "패킷"이라는 용어는 헤더 부분 및 페이로드 부분을 포함하는 데이터의 유닛에 대응할 수 있다. 문헌의 부분의 참조에 의한 임의의 통합은 또한 그 부분 내에 참조되는 용어들 또는 변수들의 정의들을 통합하도록 이해되어야 할 것이며, 그러한 정의들은 통합된 부분에서 참조된 임의의 도면들뿐만 아니라 문헌에서 어딘가에 나타난다.

[0017] [0032] 본원에서 사용되는 바와 같이, "통신 디바이스"라는 용어는 무선 통신 네트워크를 통한 음성 및/또는 데이터 통신을 위해 사용될 수 있는 전자 디바이스를 지칭한다. 통신 디바이스들의 예들은 스마트 스피커들, 스피커 바들, 셀룰러 폰들, PDA(personal digital assistant)들, 핸드헬드 디바이스들, 헤드셋들, 무선 모뎀들, 랩톱 컴퓨터들, 개인용 컴퓨터들 등을 포함한다.

[0018] [0033] 도 1은 입력 사운드를 수신하고 입력 사운드에서 하나 이상의 타깃 사운드들의 존재 또는 부재를 검출하기 위해 멀티-스테이지 타깃 사운드 검출기(120)로 입력 사운드를 프로세싱하도록 구성된 디바이스(102)를 포함하는 시스템(100)을 도시한다. 디바이스(102)는 마이크로폰(112)으로서 나타낸 하나 이상의 마이크로폰들, 및 하나 이상의 프로세서들(160)을 포함한다. 하나 이상의 프로세서들(160)은 타깃 사운드 검출기(120), 및 오디오 데이터(132)를 저장하도록 구성된 버퍼(130)를 포함한다. 타깃 사운드 검출기(120)는 제1 스테이지(140) 및 제2 스테이지(150)를 포함한다. 일부 구현들에서, 디바이스(102)는, 예시적인, 비-제한적 예들로서, 통합형 어시스턴트 애플리케이션을 갖는 무선 스피커 및 음성 커맨드 디바이스(예컨대, "스마트 스피커" 디바이스 또는 홈 오토메이션 시스템), 휴대용 통신 디바이스(예컨대, "스마트 폰" 또는 헤드셋), 또는 운송수단 시스템을 포함할 수 있다.

[0019] [0034] 마이크로폰(112)은 수신된 입력 사운드에 대한 응답으로 오디오 신호(114)를 생성하도록 구성된다. 예컨대, 입력 사운드는 타깃 사운드(106), 비-타깃 사운드(107), 또는 둘 모두를 포함할 수 있다. 오디오 신호(114)는 버퍼(130)에 제공되고, 오디오 데이터(132)로서 저장된다. 예시적인 예에서, 버퍼(130)는 PCM(pulse-code modulation) 버퍼에 대응하고, 오디오 데이터(132)는 PCM 데이터에 대응한다. 본원에서 추가로 설명되는

바와 같이, 버퍼(130)의 오디오 데이터(132)는 프로세싱을 위해 타깃 사운드 검출기(120)의 제1 스테이지(140) 및 제2 스테이지(150)에 액세스가능하다.

[0020] [0035] 타깃 사운드 검출기(120)는 오디오 신호(114)가 하나 이상의 관심 타깃 사운드들을 표시하는지 여부를 결정하기 위해 오디오 데이터(132)를 프로세싱하도록 구성된다. 예컨대, 타깃 사운드 검출기(120)는, 타깃 사운드(106)에 있을 수 있는, 알람(191), 도어벨(192), 사이렌(193), 유리 깨짐(194), 아기 울음(195), 도어 열림 또는 닫힘(196), 및 개 짖음(197)을 포함하는 일 세트의 타깃 사운드들(104) 각각을 검출하도록 구성된다. 타깃 사운드들(104)의 세트에 포함된 타깃 사운드들(191-197)은 예시적인 예들로서 제공된다는 것이 이해되어야 하며, 다른 구현들에서, 타깃 사운드들(104)의 세트는 더 적거나, 더 많거나, 또는 상이한 사운드들을 포함할 수 있다. 타깃 사운드 검출기(120)는 (비-타깃 사운드 소스(108)로서 나타낸) 하나 이상의 다른 사운드 소스들로부터 유래하는 비-타깃 사운드(107)가 타깃 사운드들(191-197) 중 어느 것도 포함하지 않는다는 것을 검출하도록 추가로 구성된다.

[0021] [0036] 타깃 사운드 검출기(120)의 제1 스테이지(140)는 오디오 데이터(132)를 프로세싱하도록 구성된 이진 타깃 사운드 분류기(144)를 포함한다. 일부 구현들에서, 이진 타깃 사운드 분류기(144)는 뉴럴 네트워크를 포함한다. 일부 예들에서, 이진 타깃 사운드 분류기(144)는 예시적인 비제한적인 예들로서 베이지안 분류기 또는 GMM(Gaussian Mixed Model) 분류기 중 적어도 하나를 포함한다. 일부 구현들에서, 이진 타깃 사운드 분류기(144)는 2개의 출력들: 분류되는 오디오 데이터(132)가 타깃 사운드들(191-197) 중 하나 이상을 포함한다는 것을 표시하는 제1 출력(예컨대, 1), 또는 오디오 데이터(132)가 타깃 사운드들(191-197) 중 어느 것도 포함하지 않음을 표시하는 제2 출력(예컨대, 0), 중 하나를 생성하도록 트레이닝된다. 예시적인 예에서, 이진 타깃 사운드 분류기(144)는 각각의 타깃 사운드들(191-197) 간에 구별하도록 트레이닝되지 않아서, 감소된 프로세싱 로드 및 더 작은 메모리 풋프린트를 가능하게 한다.

[0022] [0037] 제1 스테이지(140)는 타깃 사운드의 검출에 대한 응답으로 제2 스테이지(150)를 활성화하도록 구성된다. 예시하자면, 이진 타깃 사운드 분류기(144)는, 오디오 데이터(132)에서 다수의 타깃 사운드들(104) 중 임의의 타깃 사운드의 존재를 검출하는 것에 대한 응답으로 제2 스테이지(150)를 활성화하기 위한 신호(142)(또한 "활성화 신호"(142)로 지칭됨)를 생성하도록, 그리고 다수의 타깃 사운드들(104) 중 어느 것도 오디오 데이터(132)에 존재하지 않는다는 것을 검출하는 것에 대한 응답으로 신호(142)를 생성하는 것을 억제하도록 구성된다. 특정 양상에서, 신호(142)는 제1 값(예컨대, 제1 출력) 및 제2 값(예컨대, 제2 출력)을 포함하는 이진 신호이고, 신호(142)를 생성하는 것은 제1 값(예컨대, 로지컬 1)을 갖는 이진 신호를 생성하는 것에 대응한다. 이러한 양상에서, 신호(142)를 생성하는 것을 억제하는 것은 제2 값(예컨대, 로지컬 0)을 갖는 이진 신호를 생성하는 것에 대응한다.

[0023] [0038] 일부 구현들에서, 제2 스테이지(150)는, 도 2를 참조하여 더 설명되는 바와 같이, 신호(142)에 응답하여, 오디오 데이터(132)를 프로세싱하도록 활성화되도록 구성된다. 예시적인 예에서, 제어 레지스터의 특정 비트는 활성화 신호(142)의 존재 또는 부재를 나타내며, 제2 스테이지(150) 내의 또는 제2 스테이지(150)에 커플링된 제어 회로는 특정 비트를 판독하도록 구성된다. 비트의 "1" 값은 신호(142)를 표시하며 제2 스테이지(150)가 활성화되게 하고, 비트의 "0" 값은 신호(142)의 부재 및 오디오 데이터(132)의 현재 부분의 프로세싱의 완료 시에 제2 스테이지(150)가 비활성화될 수 있음을 표시한다. 다른 구현들에서, 예시적인 비-제한적인 예들로서, 활성화 신호(142)는 대신에 버스 또는 제어 라인 상의 디지털 또는 아날로그 신호, 인터럽트 제어기에서의 인터럽트 플래그, 또는 광학적 또는 기계적 신호로서 구현된다.

[0024] [0039] 제2 스테이지(150)는 타깃 사운드(106)의 검출에 대한 응답으로 버퍼(130)로부터 오디오 데이터(132)를 수신하도록 구성된다. 예에서, 제2 스테이지(150)는 타깃 사운드(106)를 포함하는 오디오 데이터(132)의 하나 이상의 부분들(예컨대, 프레임들)을 프로세싱하도록 구성된다. 예컨대, 버퍼(130)는 오디오 신호(114)의 일련의 프레임들을 오디오 데이터(132)로서 버퍼링할 수 있어서, 활성화 신호(142)가 생성될 때, 제2 스테이지(150)는 버퍼링된 일련의 프레임들을 프로세싱하고 검출기 출력(152)을 생성할 수 있으며, 검출기 출력(152)은, 다수의 타깃 사운드들(104) 각각에 대해, 오디오 데이터(132)에서의 그 타깃 사운드의 존재 또는 부재를 표시한다.

[0025] [0040] 비활성화될 때, 제2 스테이지(150)는 오디오 데이터(132)를 프로세싱하지 않고, 활성화될 때보다 더 적은 전력을 소비한다. 예컨대, 제2 스테이지(150)의 비활성화는, 동적 전력 소비를 감소시키기 위해, 오디오 데이터(132)가 제2 스테이지(150)에 입력되는 것을 방지하도록 제2 스테이지(150)에 대한 입력 버퍼를 게이팅하는 것, 제2 스테이지(150) 내에서 회로 스위칭을 방지하도록 클록 신호를 게이팅하는 것, 또는 둘 모두를 포함할

수 있다. 다른 예로서, 제2 스테이지(150)의 비활성화는, 회로 엘리먼트들의 상태를 잃지 않으면서 정적 전력 소비를 감소시키기 위해 제2 스테이지(150)에 대한 전력 공급을 감소시키는 것, 제2 스테이지(150)의 적어도 일부로부터 전력을 제거하는 것, 또는 이들의 조합을 포함할 수 있다.

[0026] [0041] 일부 구현들에서, 타깃 사운드 검출기(120), 버퍼(130), 제1 스테이지(140), 제2 스테이지(150), 또는 이들의 임의의 조합은 전용 회로부 또는 하드웨어를 사용하여 구현된다. 일부 구현들에서, 타깃 사운드 검출기(120), 버퍼(130), 제1 스테이지(140), 제2 스테이지(150), 또는 이들의 임의의 조합은 펌웨어 또는 소프트웨어의 실행을 통해 구현된다. 예시하자면, 디바이스(102)는 명령들을 저장하도록 구성된 메모리를 포함할 수 있고, 하나 이상의 프로세서들(160)은 타깃 사운드 검출기(120), 버퍼(130), 제1 스테이지(140), 및 제2 스테이지(150) 중 하나 이상을 구현하기 위한 명령들을 실행하도록 구성된다.

[0027] [0042] 이진 타깃 사운드 분류기(144)의 프로세싱 동작들은 제2 스테이지(150)에 의해 수행되는 프로세싱 동작들과 비교하여 덜 복잡하기 때문에, 제1 스테이지(140)에서의 오디오 데이터(132)의 올웨이즈-온 프로세싱은 제2 스테이지(150)에서 오디오 데이터(132)를 프로세싱하는 것보다 상당히 더 적은 전력을 사용한다. 그 결과, 프로세싱 자원들이 보존되고, 전체 전력 소비가 감소된다.

[0028] [0043] 일부 구현들에서, 제1 스테이지(140)는 또한 디바이스(102)의 하나 이상의 다른 컴포넌트들을 활성화하도록 구성된다. 예시적인 예에서, 도 6을 참조하여 추가 설명되는 바와 같이, 제1 스테이지(140)는 디바이스(102)의 환경(예컨대, 집에서, 실외에서, 자동차에서 등)을 검출하는 데 사용되는 카메라를 활성화하고, 제2 스테이지(150)는 검출된 환경과 연관된 타깃 사운드들에 포커싱하도록(focus on) 동작될 수 있다.

[0029] [0044] 도 2는, 이진 타깃 사운드 분류기(144)가 뉴럴 네트워크(212)를 포함하고, 이진 타깃 사운드 분류기(144) 및 버퍼(130)가 저전력 도메인(203), 이를테면 하나 이상의 프로세서들(160)의 올웨이즈-온 저전력 도메인에 포함되는, 디바이스(102)의 예(200)를 도시한다. 제2 스테이지(150)는 온-디멘드 전력 도메인과 같은 다른 전력 도메인(205)에 있다. 일부 구현들에서, 타깃 사운드 검출기(120)의 제1 스테이지(140)(예컨대, 이진 타깃 사운드 분류기(144)) 및 버퍼(130)는 올웨이즈-온 모드로 동작하도록 구성되고, 타깃 사운드 검출기(120)의 제2 스테이지(150)는 온-디멘드 모드로 동작하도록 구성된다.

[0030] [0045] 전력 도메인(205)은 타깃 사운드 검출기(102)의 제2 스테이지(150), 사운드 콘텍스트 애플리케이션(240), 및 활성화 회로부(230)를 포함한다. 활성화 회로부(230)는 제2 스테이지(150)와 같은 전력 도메인(205)의 하나 이상의 컴포넌트들을 선택적으로 활성화하기 위해 활성화 신호(142)(예컨대, 웨이크업 인터럽트 신호)에 응답한다. 예시하자면, 일부 구현들에서, 활성화 회로부(230)는 신호(142)의 수신에 응답하여 제2 스테이지(150)를 저전력 상태(232)로부터 활성 상태(234)로 전환하도록 구성된다.

[0031] [0046] 예컨대, 활성화 회로부(230)는 전력 관리 회로부, 클록 회로부, 헤드 스위치 또는 풋 스위치 회로부, 버퍼 제어 회로부, 또는 이들의 임의의 조합을 포함하거나 그에 커플링될 수 있다. 활성화 회로부(230)는, 이를테면, 제2 스테이지(150)의 전력 공급부의 전압, 전력 도메인(205)의 전압, 또는 둘 모두의 전압을 선택적으로 인가 또는 상승시킴으로써, 제2 스테이지(150)의 파워-온(powering-on)을 개시하도록 구성될 수 있다. 다른 예로서, 활성화 회로부(230)는 이를테면, 전력 공급부를 제거하지 않고 회로 동작을 방지하기 위해 제2 스테이지(150)에 대한 클록 신호를 선택적으로 게이팅 또는 언-게이팅하도록 구성될 수 있다.

[0032] [0047] 제2 스테이지(150)는, 다수의 타깃 사운드들(104) 각각에 대해, 오디오 데이터(132)에서 그 타깃 사운드의 존재 또는 부재를 표시하는 검출기 출력(152)을 생성하도록 구성된 다중 타깃 사운드 분류기(210)를 포함한다. 다수의 타깃 사운드들은 사운드 이벤트들의 다수의 클래스들(290)에 대응하며, 사운드 이벤트들의 다수의 클래스들(290)은, 알람(291), 도어벨(292), 사이렌(293), 유리 깨짐(294), 아기 울음(295), 도어 열림 또는 닫힘(296), 또는 개 짖음(297) 중 적어도 2개를 포함한다. 사운드 이벤트 클래스들(291-297)은 예시적인 예들로서 제공된다는 것이 이해되어야 한다. 다른 예들에서, 다수의 클래스들(290)은 더 적은, 더 많은, 또는 상이한 사운드 이벤트들을 포함한다. 예컨대, 디바이스(102)가 운송수단(예컨대, 자동차)에서 구현되는 구현에서, 다수의 클래스들(290)은, 예시적인 비제한적인 예들로서, 운송수단 도어 열림 또는 닫힘, 도로 소음, 창문 열림 또는 닫힘, 라디오, 제동, 핸드 브레이크 맞물림 또는 맞물림해제, 앞유리 와이퍼(windshield wiper)들, 방향 지시등, 또는 엔진 회전 중 하나 이상과 같이, 운송수단에서 보다 일반적으로 직면하는 사운드 이벤트들을 포함한다. 사운드 이벤트 클래스들(예컨대, 다수의 클래스들(290))의 단일 세트가 도시되지만, 다른 구현들에서, 도 6을 참조로 더 설명되는 바와 같이, 다중 타깃 사운드 분류기(210)는 디바이스(102)의 환경(예컨대, 디바이스(102)가 집에 있을 때 타깃 사운드들의 일 세트, 및 디바이스(102)가 운송수단에 있을 때 타깃 사운드들의 다른 세트)에 기반하여 사운드 이벤트 클래스들의 다수의 세트들 사이로부터 선택하도록 구성된다.

[0033] [0048] 일부 구현들에서, 다중 타깃 사운드 분류기(210)는 오디오 데이터(132)의 "실시간보다 더 빠른" 프로세싱을 수행한다. 예시적인 비제한적인 예에서, 버퍼(130)는, 버퍼(130) 내의 가장 오래된 오디오 데이터가 가장 최근에 수신된 오디오 데이터로 대체되는 원형 버퍼 구성에서, 대략 2초의 오디오 데이터를 저장하도록 사이즈가 정해진다. 제1 스테이지(140)는 순차적으로 수신된 오디오 데이터(132)의 20 밀리초(mS) 세그먼트들(예컨대, 프레임들)을, 실시간 방식으로(예컨대, 이진 타깃 사운드 분류기(144)는 매 20 mS마다 하나의 20 mS 세그먼트를 프로세싱함) 그리고 저전력 소비로, 주기적으로 프로세싱하도록 구성될 수 있다. 그러나, 제2 스테이지(150)가 활성화될 때, 다중 타깃 사운드 분류기(210)는, 버퍼링된 오디오 데이터(132)를 더 신속하게 프로세싱하여 검출기 출력(152)을 생성하기 위해, 더 빠른 레이트 및 더 높은 전력 소비로, 버퍼링된 오디오 데이터(132)를 프로세싱한다.

[0034] [0049] 일부 구현들에서, 검출기 출력(152)은 그 타깃 사운드의 검출(또는 검출의 가능성)을 나타내는 다수의 값들, 이를테면 각각의 타깃 사운드에 대한 비트 또는 멀티-비트 값을 포함한다. 예시적인 예에서, 검출기 출력(152)은 7-비트 값을 포함하며, 제1 비트는 알람(291)으로 분류된 사운드의 검출 또는 비-검출에 대응하고, 제2 비트는 도어벨(292)로 분류된 사운드의 검출 또는 비-검출에 대응하고, 제3 비트는 사이렌(293)으로 분류된 사운드의 검출 또는 비-검출에 대응하고, 제4 비트는 유리 깨짐(294)으로 분류된 사운드의 검출 또는 비-검출에 대응하고, 제5 비트는 아기 울음(295)으로 분류된 사운드의 검출 또는 비-검출에 대응하고, 제6 비트는 도어 열림 또는 닫힘(296)으로 분류된 사운드의 검출 또는 비-검출에 대응하고, 그리고 제7 비트는 개 짖음(297)으로 분류된 사운드의 검출 또는 비-검출에 대응한다.

[0035] [0050] 제2 스테이지(150)에 의해 생성된 검출기 출력(152)은 사운드 콘텍스트 애플리케이션(240)에 제공된다. 사운드 콘텍스트 애플리케이션(240)은 하나 이상의 타깃 사운드들의 검출에 기반하여 하나 이상의 동작들을 수행하도록 구성될 수 있다. 예시하자면, 디바이스(102)가 홈 오토메이션 시스템 내에 있는 구현에서, 사운드 콘텍스트 애플리케이션(240)은 하나 이상의 검출된 사운드 이벤트들을 사용자에게 경고하기 위해 사용자 인터페이스 신호(242)를 생성할 수 있다. 예컨대, 사용자 인터페이스 신호(242)는 출력 디바이스(250)(예컨대, 스피치 인터페이스 디바이스의 디스플레이 스크린 또는 라우드스피커)로 하여금 개 짖음 및 유리 깨짐이 건물의 뒷문에서 검출되었음을 사용자에게 경고하게 할 수 있다. 다른 예에서, 사용자가 건물 내에 있지 않을 때, 사용자 인터페이스 신호(242)는 출력 디바이스(250)(예컨대, 셀룰러 네트워크 또는 무선 근거리 네트워크와 같은 무선 네트워크에 커플링된 송신기)로 하여금 사용자의 폰 또는 스마트 워치에 경고를 송신하게 할 수 있다.

[0036] [0051] 디바이스(102)가 운송수단(예컨대, 자동차) 내에 있는 다른 구현에서, 사운드 콘텍스트 애플리케이션(240)은, 출력 디바이스(250)(예컨대, 디스플레이 스크린 또는 음성 인터페이스)를 통해, 운송수단이 움직이는 동안 외부 마이크로폰을 통해 사이렌이 검출되었음을 운송수단의 조작자에게 경고하기 위해 사용자 인터페이스 신호(242)를 생성할 수 있다. 운송수단이 턴 오프되고 조작자가 운송수단에서 하차하면, 사운드 콘텍스트 애플리케이션(240)은, 운송수단의 내부 마이크로폰을 통해 우는 아기가 검출되었음을, 출력 디바이스(250)를 통해(예컨대, 소유자의 폰 또는 스마트 워치로의 무선 송신), 운송수단의 소유자에게 경고하기 위해 사용자 인터페이스 신호(242)를 생성할 수 있다.

[0037] [0052] 예시적인 예들로서, 디바이스(102)가 헤드폰들 또는 헤드셋과 같은 오디오 재생 디바이스에 통합되거나 그에 커플링되는 다른 구현에서, 사운드 콘텍스트 애플리케이션(240)은, 출력 디바이스(250)(예컨대, 디스플레이 스크린 또는 라우드스피커)을 통해, 사이렌이 검출되었음을 재생 디바이스(playback device)의 사용자에게 경고하기 위해 사용자 인터페이스 신호(242)를 생성할 수 있거나, 또는 헤드폰들 또는 헤드셋의 라우드스피커에서의 재생을 위해 사이렌을 통과(pass-through)할 수 있다.

[0038] [0053] 전력 도메인(205)에서 활성화 회로부(230)가 제2 스테이지(150)와 별개인 것으로 예시되지만, 다른 구현들에서, 활성화 회로부(230)는 제2 스테이지(150)에 포함될 수 있다. 일부 구현들에서, 출력 디바이스(250)는 디스플레이 스크린 또는 라우드스피커와 같은, 디바이스(102)의 사용자 인터페이스 컴포넌트로서 구현되지만, 다른 구현들에서 출력 디바이스(250)는 디바이스(102)로부터 원격에 있고 디바이스(102)에 커플링되는 사용자 인터페이스 디바이스일 수 있다. 다중 타깃 사운드 분류기(210)가 7개의 클래스들(291-297)에 대응하는 사운드 이벤트들을 검출하고 이들 사이를 구별하도록 구성되지만, 다른 구현들에서 다중 타깃 사운드 분류기(210)는 7개의 클래스들(291-297) 중 임의의 하나 이상의 클래스 대신에, 또는 추가로, 임의의 다른 사운드 이벤트를 검출하도록 구성될 수 있고, 다중 타깃 사운드 분류기(210)는 임의의 다른 개수의 클래스들에 따라 사운드 이벤트들을 분류하도록 구성될 수 있다.

[0039] [0054] 도 3은, 디바이스(102)가 버퍼(130) 및 타깃 사운드 검출기(120)를 포함하고 또한 오디오 장면 검출기

(302)를 포함하는 구현(300)을 도시한다. 오디오 장면 검출기(302)는 오디오 장면 변화 검출기(304) 및 오디오 장면 분류기(308)를 포함한다. 오디오 장면 변화 검출기(304)는 오디오 데이터(132)를 프로세싱하도록 그리고 오디오 장면 변화의 검출에 대한 응답으로 장면 변화 신호(306)를 생성하도록 구성된다. 일부 구현들에서, 오디오 장면 변화 검출기(304)는 오디오 장면 검출기(302)의 제1 스테이지(예컨대, 저전력, 올웨이즈-온(always-on) 프로세싱 스테이지)에서 구현되고, 오디오 장면 분류기(308)는, 도 2의 다중 타깃 사운드 분류기(210)가 활성화 신호(142)에 의해 활성화되는 것과 유사한 방식으로 장면 변화 신호(306)에 의해 활성화되는 오디오 장면 검출기(302)의 제2 스테이지(예컨대, 더 강력한 고성능 프로세싱 스테이지)에서 구현된다. 타깃 사운드 검출과 달리, 오디오 환경이 항상 존재하고, 오디오 장면 검출기(302)의 동작의 효율성은 제1 스테이지에서, 정확한 오디오 환경을 식별하는 것과 연관된 계산적 패널티를 발생시키지 않으면서 오디오 환경에서의 변화들을 검출함으로써 향상된다.

[0040] [0055] 일부 구현들에서, 오디오 장면 변화 검출기(304)는 소음 통계(noise statistics)(310) 또는 비-정지형(non-stationary) 사운드 통계(312) 중 적어도 하나에서의 변화들을 검출하는 것에 기반하여 오디오 장면에서의 변화를 검출하도록 구성된다. 예로서, 오디오 장면 변화 검출기(304)는, 비교적 큰 시간 윈도우(예컨대, 3-5초)에 걸쳐 시간-평균된, 소음 통계(310)(예컨대, 소음을 포함하는 것으로 식별된 오디오 프레임들의 평균 스펙트럼 에너지 분포) 및 비-정지형 사운드 통계(312)(예컨대, 비-정지형 사운드를 포함하는 것으로 식별된 오디오 프레임들의 평균 스펙트럼 에너지 분포)를 결정하기 위해 오디오 데이터(132)를 프로세싱한다. 오디오 장면들 사이의 변화들은, 소음 통계(310), 비-정지형 사운드 통계(312), 또는 둘 모두의 변화를 결정하는 것에 기반하여 검출된다. 예컨대, 사무실 환경의 소음 및 사운드 특징들은 이동하는 자동차 내의 소음 및 사운드 특징들과 충분히 구별되어, 사무실 환경으로부터 운송수단 환경으로의 변화가 검출될 수 있고, 일부 구현들에서는, 소음 및 사운드 특성들을 사무실 환경 또는 운송수단 환경 중 어느 하나에 대응하는 것으로서 식별하지 않고 변화가 검출된다. 오디오 장면 변화를 검출하는 것에 대한 응답으로, 오디오 장면 변화 검출기는 장면 변화 신호(306)를 생성하여 오디오 장면 분류기(308)에 전송한다.

[0041] [0056] 오디오 장면 분류기(308)는 오디오 장면 변화의 검출에 대한 응답으로 버퍼(130)로부터 오디오 데이터(132)를 수신하도록 구성된다. 일부 구현들에서, 오디오 장면 분류기(308)는 오디오 장면 변화 검출기(304)보다 더 강력하고 더 높은 복잡성의 프로세싱 컴포넌트이며, 다수의 오디오 장면 클래스들(330) 중 특정한 하나의 클래스에 대응하는 것으로 오디오 데이터(132)를 분류하도록 구성된다. 일 예에서, 다수의 오디오 장면 클래스들(330)은 집에서(332), 사무실에서(334), 식당에서(336), 자동차에서(338), 기차에서(340), 거리에서(342), 실내에서(344), 및 실외에서(346)를 포함한다.

[0042] [0057] 장면 검출기 출력(352)은 오디오 장면 검출기(302)에 의해 생성되고, 검출된 오디오 장면의 표시를 제시하며, 이는 도 2의 사운드 콘텍스트 애플리케이션(240)에 제공될 수 있다. 예컨대, 사운드 콘텍스트 애플리케이션(240)은 검출된 오디오 장면에 기반하여 디바이스(102)의 동작을 조정할 수 있는데, 이를테면, 환경과 연관된 최상위-레벨 메뉴 아이템들을 제시하기 위해 디스플레이 스크린에서 그래픽 사용자 인터페이스(GUI)를 변경할 수 있다. 예시하자면, 예시적인 비-제한적인 예들로서, 검출된 환경이 자동차에서 일 때 내비게이션 및 통신 아이템들(예컨대, 핸즈-프리 다이얼링)이 제공될 수 있고, 검출된 환경이 실외에서 일 때 카메라 및 오디오 레코딩 아이템들이 제공될 수 있으며, 그리고 검출된 환경이 사무실에서 일 때 메모 작성 및 연락처 아이템들이 제공될 수 있다.

[0043] [0058] 다수의 오디오 장면 클래스들(330)이 8개의 클래스들(332-346)을 포함하는 것으로 설명되지만, 다른 구현들에서 다수의 오디오 장면 클래스들(330)은 집에서(332), 사무실(334)에서, 식당에서(336), 자동차에서(338), 기차에서(340), 거리에서(342), 실내에서(344), 또는 실외에서(346) 중 적어도 2개를 포함할 수 있다. 다른 구현들에서, 클래스들(330) 중 하나 이상은 생략될 수 있거나, 하나 이상의 다른 클래스들이 클래스들(332-346) 대신에 또는 그에 부가하여 사용될 수 있거나, 또는 이들의 임의의 조합이 사용될 수 있다.

[0044] [0059] 도 4는 오디오 장면 변화 검출기(304)가 장면들 사이의 전환들에 대응하는 오디오 데이터를 사용하여 트레이닝되는 장면 전환 분류기(414)를 포함하는 오디오 장면 변화 검출기(304)의 구현(400)을 도시한다. 예컨대, 장면 전환 분류기(414)는 사무실에서 거리로의 전환, 자동차에서 실외로의 전환, 식당에서 거리로의 전환 등에 대해 캡처된 오디오 데이터에 대해 트레이닝될 수 있다. 일부 구현들에서, 장면 전환 분류기(414)는 도 3을 참조하여 설명된 오디오 장면 변화 검출기(304)의 구현보다 더 작은 모델을 사용하는 더 강건한 변화 검출을 제공한다.

[0045] [0060] 도 5는, 오디오 장면 변화 검출기(304)가 오디오 장면 분류기(308)와 비교하여 감소된 오디오 장면들의

세트를 사용하여 오디오 데이터(132)를 분류하도록, 오디오 장면 검출기(302)가 계층적 검출기에 대응하는 구현(500)을 도시한다. 예시하자면, 오디오 장면 변화 검출기(304)는 감소된 클래스들의 세트(530)의 오디오 장면 클래스들 사이의 변화들을 검출하는 것에 기반하여 오디오 장면 변화를 검출하도록 구성된 계층적 모델 변화 검출기(514)를 포함한다. 예컨대, 감소된 클래스들의 세트(530)는 "운송수단에서" 클래스(502), 실내에서 클래스(344) 및 실외에서 클래스(346)를 포함한다. 일부 구현들에서, 감소된 클래스들의 세트(530) 중 하나 이상(또는 전부)은 오디오 장면 분류기(308)에 의해 사용되는 다수의 클래스들을 포함하거나 또는 그에 걸쳐 있다. 예시하자면, "운송수단에서" 클래스(502)는 오디오 장면 분류기(308)가 "자동차에서" 또는 "기차에서"로서 구별하는 오디오 장면들을 분류하는 데 사용된다. 일부 구현들에서, 감소된 클래스들의 세트(530) 중 하나 이상(또는 전부)은 오디오 장면 분류기(308)에 의해 사용되는 클래스들(330)의 서브세트, 이를테면 실내에서 클래스(344) 및 실외에서 클래스(346)를 형성한다. 일부 예들에서, 감소된 클래스들의 세트(530)는, 오디오 장면 변화들을 검출할 개선된 확률을 위해, 가장 가능성이 높은 조우되는 오디오 장면들 중 2개 또는 3개를 포함하도록 구성된다.

[0046] [0061] 감소된 클래스들의 세트(530)는 오디오 장면 분류기(308)의 클래스들(330)과 비교하여 감소된 수의 클래스들을 포함한다. 예시하자면, 감소된 클래스들의 세트(530)의 오디오 장면 클래스들의 제1 카운트(3)는 오디오 장면 클래스들(330)의 제2 카운트(8) 미만이다. 감소된 클래스들의 세트(530)가 3개의 클래스들을 포함하는 것으로 설명되지만, 다른 구현들에서 감소된 클래스들의 세트(530)는 오디오 장면 분류기(308)에 의해 지원되는 클래스들의 수보다 더 적은 임의의 수의 클래스들(예컨대, 3, 4, 또는 그 초과의 클래스들과 같은 적어도 2개의 클래스들)을 포함할 수 있다.

[0047] [0062] 계층적 모델 변화 검출기(514)는 오디오 장면 분류기(308)와 비교하여 더 작은 클래스들의 세트 중으로부터 검출을 수행하기 때문에, 오디오 장면 변화 검출기(304)는 더 강력한 오디오 장면 분류기(308)에 비해 감소된 복잡성 및 전력 소비로 장면 변화들을 검출할 수 있다. 이를테면, 운송수단 또는 실외 환경으로의 중간 전환 없이 "집에서"로부터 "레스토랑에서"로(예컨대, 둘 모두는 "실내" 클래스(344)에 있음) 직접 전환하는 것과 같은, 계층적 모델 변화 검출기(514)에 의해 검출되지 않는 환경들 간의 전환들은, 발생할 가능성이 없을 수 있다.

[0048] [0063] 도 3 내지 도 5는 오디오 장면 검출기(304) 및 타깃 사운드 검출기(120) 둘 모두가 디바이스(102)에 포함되는 다양한 구현들에 대응하지만, 다른 구현들에서 오디오 장면 검출기(302)는 타깃 사운드 검출기를 포함하지 않는 디바이스에 구현될 수 있다. 예시적인 예에서, 디바이스(102)는 버퍼(130) 및 오디오 장면 검출기(302)를 포함하고, 타깃 사운드 검출기(120)의 제1 스테이지(140), 제2 스테이지(150) 또는 둘 모두를 생략한다.

[0049] [0064] 도 6은 디바이스(102)가 카메라, 위치 검출 시스템, 또는 오디오 장면 검출기 중 적어도 하나에 기반하여 환경을 검출하도록 구성된 장면 검출기(606)를 포함하는 특정한 예(600)를 도시한다.

[0050] [0065] 디바이스(102)는 환경(608)을 결정하는 데 장면 검출기(606)에 의해 사용가능한 데이터를 생성하는 하나 이상의 센서들(602)을 포함한다. 하나 이상의 센서들(602)은 카메라(620) 및 GPS(global positioning system) 수신기(624)로서 각각 예시된, 위치 검출 시스템의 하나 이상의 센서들 및 하나 이상의 카메라들을 포함한다. 카메라(620)는 임의의 타입의 이미지 캡처 디바이스를 포함할 수 있고, 정지 이미지 또는 비디오 캡처, 가시, 적외선 또는 자외선 스펙트럼들, 깊이 감지(예컨대, 구조화된 광, 비행 시간(time-of-flight)), 임의의 다른 이미지 캡처 기법, 또는 이들의 임의의 조합을 지원하거나 포함할 수 있다.

[0051] [0066] 제1 스테이지(140)는 제1 스테이지(140)에 의한 타깃 사운드의 검출에 대한 응답으로 저전력 상태로부터 센서들(602) 중 하나 이상을 활성화하도록 구성된다. 예컨대, 신호(142)는 카메라(620) 및 GPS 수신기(624)에 제공될 수 있다. 카메라(620) 및 GPS 수신기(624)는 (예컨대, 디바이스(102)의 다른 애플리케이션에 의해 사용되지 않을 때) 저전력 상태로부터 활성 상태로 전환하도록 신호(142)에 응답한다.

[0052] [0067] 장면 검출기(606)는 오디오 장면 검출기(302)를 포함하고, 카메라(620), GPS 수신기(624) 또는 오디오 장면 검출기(302) 중 적어도 하나에 기반하여 환경(608)을 검출하도록 구성된다. 제1 예로서, 장면 검출기(606)는 카메라(624)로부터의 입력 신호(622)(예컨대, 이미지 데이터)에 적어도 부분적으로 기반하여 디바이스(102)의 환경(608)의 제1 추정치를 생성하도록 구성된다. 예시하자면, 장면 검출기(606)는 시각적 특징들에 기반하여, 집에서, 사무실에서, 식당에서, 자동차에서, 기차에서, 거리에서, 실외에서, 또는 실내에서와 같은 환경(608)의 제1 분류를 생성하기 위해 입력 데이터(622)를 프로세싱하도록 구성될 수 있다.

[0053] [0068] 제2 예로서, 장면 검출기(606)는 GPS 수신기로부터의 위치 정보(626)에 적어도 부분적으로 기반하여 환경(608)의 제2 추정치를 생성하도록 구성된다. 예시하자면, 장면 검출기(606)는 위치가 사용자의 집, 사용자의 사무실, 식당, 기차 루트, 거리, 실외 위치, 또는 실내 위치에 대응하는지 여부를 결정하기 위해 위치 정보(626)를 사용하여 맵 데이터를 검색할 수 있다. 장면 검출기(606)는 디바이스(102)가 자동차에서 이동하고 있는지 아니면 비행기에서 이동하고 있는지를 결정하기 위해 위치 데이터(626)에 기반하여 디바이스(102)의 이동 속력을 결정하도록 구성될 수 있다.

[0054] [0069] 일부 구현들에서, 장면 검출기(606)는 제1 추정치, 제2 추정치, 오디오 장면 검출기(302)의 장면 검출기 출력(352), 및 제1 추정치, 제2 추정치, 및 장면 검출기 출력(352)과 연관된 개개의 신뢰 레벨들에 기반하여 환경(608)을 결정하도록 구성된다. 환경(608)의 표시가 타깃 사운드 검출기(120)에 제공되며, 다중 타깃 사운드 분류기(210)의 동작은 장면 검출기(606)에 의한 환경(608)의 분류에 적어도 부분적으로 기반한다.

[0055] [0070] 도 6은 카메라(620), GPS 수신기(624) 및 오디오 장면 검출기(302)를 포함하는 디바이스(102)를 도시하지만, 다른 구현들에서, 카메라(620), GPS 수신기(624) 또는 오디오 장면 검출기(302) 중 하나 이상이 생략되거나, 하나 이상의 다른 센서들이 추가되거나, 또는 이들의 임의의 조합이 이용된다. 예컨대, 오디오 장면 검출기(302)는 하나 이상의 다른 오디오 장면 검출기들로 대체되거나 생략될 수 있다. 다른 예들에서, 장면 검출기(606)는 카메라로부터의 이미지 데이터(622)에만 전적으로 기반하여, GPS 센서(624)로부터의 위치 데이터(62)에만 전적으로 기반하여, 또는 오디오 장면 검출기로부터의 장면 검출에만 전적으로 기반하여 환경(608)을 결정한다.

[0056] [0071] 하나 이상의 센서들(602), 오디오 장면 검출기(302), 및 장면 검출기(606)는 신호(142)에 응답하여 활성화되지만, 다른 구현들에서 장면 검출기(606), 오디오 장면 검출기(302), 센서들(602) 중 하나 이상 또는 이들의 임의의 조합은 신호(142)와 독립적으로 활성화 또는 비활성화될 수 있다. 비-제한적인 예로서, 운송수단 또는 가전제품에서와 같은 비-전력-제약 환경에서, 하나 이상의 센서들(602), 오디오 장면 검출기(302), 및 장면 검출기(606)는, 타깃 사운드 활동이 전혀 검출되지 않는다하더라도, 활성 상태를 유지할 수 있다.

[0057] [0072] 도 7은 다중 타깃 사운드 분류기(210)가 환경(608)에 대응하는 사운드 이벤트들의 다수의 클래스들(290) 중 하나 이상의 특정 클래스들(702)에 포커싱하도록 조정되는 예(700)를 도시한다. 예(700)에서, 환경(608)은 "자동차에서"로 검출되고, 다중 타깃 사운드 분류기(210)는, 오디오 데이터(132)에서의 타깃 사운드를 자동차에서 보다 일반적으로 직면하게 되는 다수의 클래스들(290) 중의 클래스들: 사이렌(293), 유리 깨짐(294), 아기 울음(295), 또는 도어 열림 또는 닫힘(296) 중 하나로서 식별하는데 더 포커싱하고, 그리고 타깃 사운드를 자동차에서 일반적으로 덜 직면하게 되는 클래스들: 알람(291), 도어벨(292) 또는 개 짖음(297) 중 하나로서 식별하는 데 덜 포커싱하도록 조정된다. 결과적으로, 타깃 사운드 검출은 타깃 사운드 검출을 포커싱하기 위해 환경 정보가 사용되지 않는 구현들에서보다 더 정확하게 수행될 수 있다.

[0058] [0073] 도 8은 다중 타깃 사운드 분류기(210)가 사운드 이벤트 클래스들의 다수의 세트들 중에서 환경(608)에 대응하는 사운드 이벤트 클래스들의 특정 세트를 선택하도록 구성되는 예(800)를 도시한다. 트레이닝된 데이터의 제1 세트(802)는 제1 환경(예컨대, 집에서)과 연관된 사운드 이벤트 클래스들(812)의 제1 세트를 포함한다. 트레이닝된 데이터의 제2 세트(804)는 제2 환경(예컨대, 자동차에서)과 연관된 사운드 이벤트 클래스들의 제2 세트(814)를 포함하고, 그리고 트레이닝된 데이터의 하나 이상의 추가적인 세트들은, 제N 환경(예컨대, 사무실에서)과 연관된 사운드 이벤트 클래스들의 제N 세트(818)를 포함하는 트레이닝된 데이터의 제N 세트(808)를 포함하며, N은 1보다 큰 정수이다. 비-제한적인 예에서, 트레이닝된 데이터(802-808)의 세트들 각각은 클래스들(330) 중 하나(예컨대, N=8)에 대응한다. 일부 구현들에서, 트레이닝된 데이터의 세트들(802-808) 중 하나 이상은 환경이 결정되지 않을 때 사용될 트레이닝된 데이터의 디폴트 세트에 대응한다. 예로서, 도 2의 다수의 클래스들(290)이 트레이닝된 데이터의 디폴트 세트로서 사용될 수 있다.

[0059] [0074] 예시적인 구현에서, 사운드 이벤트 클래스들의 제1 세트(812)는 "집에서"에 대응하고, 사운드 이벤트 클래스들의 제2 세트(814)는 "자동차에서"에 대응한다. 사운드 이벤트 클래스들(812)의 제1 세트는 보다 보편적으로 집에서 직면하는 사운드 이벤트들, 이를테면 예시적인 비제한적인 예들로서, 화재 경보기, 아기 울음, 개 짖음, 도어벨, 도어 열림 또는 닫힘, 및 유리 깨짐 중 하나 이상을 포함할 수 있다. 이벤트 클래스들의 제2 세트(814)는, 예시적인 비제한적인 예들로서, 자동차 도어 열림 또는 닫힘, 도로 소음, 창문 열림 또는 닫힘, 라디오, 제동, 핸드 브레이크 맞물림 또는 맞물림해제, 앞유리 와이퍼들, 방향 지시등, 또는 엔진 회전 중 하나 이상과 같이, 자동차에서 보다 일반적으로 직면하는 사운드 이벤트들을 포함한다. 환경(608)이 "집에서"로서 검출되는 것에 대한 응답으로, 다중 타깃 사운드 분류기(210)는 그 특정 세트(즉, 사운드 이벤트 클래스들의 제

1 세트(812))의 사운드 이벤트 클래스들에 기반하여 오디오 데이터(132)를 분류하기 위해 사운드 이벤트 클래스들의 제1 세트(812)를 선택한다. 환경(608)이 "자동차에서"로서 검출되는 것에 대한 응답으로, 다중 타깃 사운드 분류기(210)는 그 특정 세트(즉, 사운드 이벤트 클래스들의 제2 세트(814))의 사운드 이벤트 클래스들에 기반하여 오디오 데이터(132)를 분류하기 위해 사운드 이벤트 클래스들의 제2 세트(814)를 선택한다.

[0060] [0075] 결과적으로, 임의의 특정 환경에 대한 타깃 사운드 분류를 수행하기 위한 전체 프로세싱 및 메모리 사용량을 증가시키지 않으면서, 각각의 환경에 대해 사운드 이벤트들의 상이한 세트들을 사용함으로써, 더 많은 전체 수의 타깃 사운드들이 검출될 수 있다. 또한, 센서들(602), 장면 검출기(606) 또는 둘 모두를 활성화하기 위해 제1 스테이지(140)를 사용함으로써, 센서들(602) 및 장면 검출기(606)의 올웨이즈-온 동작과 비교하여 전력 소비가 감소된다.

[0061] [0076] 예(800)는 환경(608)에 기반하여 사운드 이벤트 클래스들(812-818)의 세트들 중 하나를 선택하는 것으로 다중 타깃 사운드 분류기(210)를 설명하지만, 일부 구현들에서 트레이닝된 데이터(802-808)의 세트들 각각은 또한 특정 환경과 연관된 타깃 사운드들의 존재 또는 부재를 그룹으로서 검출하기 위해 이진 타깃 사운드 분류기(144)에 대한 트레이닝된 데이터를 포함한다. 예에서, 타깃 사운드 검출기(120)는 트레이닝된 데이터(802-808)의 세트들 중에서, 디바이스(102)의 검출된 환경(608)에 대응하는 트레이닝된 데이터의 특정 세트를 선택하고, 트레이닝된 데이터의 특정 세트에 기반하여 오디오 데이터(132)를 프로세싱하도록 구성된다.

[0062] [0077] 도 9는 하나 이상의 프로세서들(160)을 포함하는 집적 회로(902)로서의 디바이스(102)의 구현(900)을 도시한다. 집적 회로(902)는 또한 오디오 신호(114)가 마이크로폰(112)으로부터 수신될 수 있게 하기 위한 센서 신호 입력(910), 이를테면 하나 이상의 제1 버스 인터페이스들을 포함한다. 예컨대, 센서 신호 입력(910)은 마이크로폰(112)으로부터 오디오 신호(114)를 수신하고 오디오 신호(114)를 버퍼(130)에 제공한다. 집적 회로(902)는 또한 검출기 출력(152)을 (예컨대, 예시적인 비제한적인 예들로서 디스플레이 디바이스, 메모리 또는 송신기에) 전송하는 것을 가능하게 하기 위해 데이터 출력(912), 이를테면 제2 버스 인터페이스를 포함한다. 예컨대, 타깃 사운드 검출기(120)는 검출기 출력(152)을 데이터 출력(912)에 제공하고, 데이터 출력(912)은 검출기 출력(152)을 전송한다. 집적 회로(902)는 멀티-스테이지 타깃 사운드 검출의 구현을, 하나 이상의 마이크로폰들을 포함하는 시스템, 이를테면, 도 10 또는 도 11에 도시된 바와 같은 운송수단, 도 12에 도시된 바와 같은 가상 현실 또는 증강 현실 헤드셋, 도 13에 도시된 바와 같은 웨어러블 전자 디바이스, 도 14에 도시된 바와 같은 음성-제어 스피커 시스템, 또는 도 16에 도시된 바와 같은 무선 통신 디바이스 내의 컴포넌트로서 가능하게 한다.

[0063] [0078] 도 10은 디바이스(102)가 자동차로 예시된 운송수단(1002)에 대응하거나 또는 운송수단(1002) 내에 통합되는 구현(1000)을 도시한다. 일부 구현들에서, 멀티-스테이지 타깃 사운드 검출은 이를테면, 자동차에서 우는 아기에 대해 내부 마이크로폰들로부터 수신된 오디오 신호에 기반하여, 사이렌, 또는 둘 모두에 대한 것과 같은 외부 마이크로폰들(예컨대, 마이크로폰(112))로부터 수신된 오디오 신호에 기반하여 수행될 수 있다. 도 1의 검출기 출력(152)은 운송수단(1002)의 디스플레이 스크린에, 사용자의 모바일 디바이스에, 또는 둘 모두에 제공될 수 있다. 예컨대, 출력 디바이스(250)는 타깃 사운드(예컨대, 사이렌)가 운송수단(1002) 외부에서 검출되었음을 나타내는 통지를 디스플레이하는 디스플레이 스크린을 포함한다. 다른 예로서, 출력 디바이스(250)는 타깃 사운드(예컨대, 아기의 울음)가 운송수단(1002)에서 검출되었음을 나타내는 통지를 모바일 디바이스에 송신하는 송신기를 포함한다.

[0064] [0079] 도 11은 디바이스(102)가 유인 또는 무인 항공 디바이스(예컨대, 패키지 배달 드론)로서 예시된, 운송수단(1102)에 대응하거나 운송수단(1102) 내에 통합되는 다른 구현(1100)을 도시한다. 멀티-스테이지 타깃 사운드 검출은 이를테면, 도어의 열림 또는 닫힘에 대한 것과 같은 운송수단(1102)의 하나 이상의 마이크로폰들(예컨대, 마이크로폰(112))로부터 수신된 오디오 신호에 기반하여 수행될 수 있다. 예로서, 출력 디바이스(250)는 운송수단(1102)에 의해 타깃 사운드(예컨대, 도어의 열림 또는 닫힘)가 검출되었음을 나타내는 통지를 제어 디바이스에 송신하는 송신기를 포함한다.

[0065] [0080] 도 12는 디바이스(102)가 가상 현실, 증강 현실 또는 혼합 현실 헤드셋(1202)에 대응하는 휴대용 전자 디바이스인 구현(1200)을 도시한다. 하나 이상의 프로세서들(160) 및 마이크로폰(112)은 헤드셋(1202)에 통합된다. 멀티-스테이지 타깃 사운드 검출은 헤드셋(1202)의 마이크로폰(112)으로부터 수신된 오디오 신호에 기반하여 수행될 수 있다. 시각적 인터페이스 디바이스, 이를테면 출력 디바이스(250)는, 헤드셋(1202)이 착용된 동안 사용자에게 증강 현실 또는 가상 현실 이미지들 또는 장면들의 디스플레이를 가능하게 하기 위해 사용자의 눈 앞에 포지셔닝된다. 특정 예에서, 출력 디바이스(250)는 타깃 사운드(예컨대, 화재 경보 또는 도어벨)가 헤

드셋(1202) 외부에서 검출됨을 나타내는 통지를 디스플레이하도록 구성된다.

[0066] [0081] 도 13은 디바이스(102)가 "스마트 워치"로서 예시된 웨어러블 전자 디바이스(1302)에 대응하는 휴대용 전자 디바이스인 구현(1300)을 도시한다. 하나 이상의 프로세서들(160) 및 마이크로폰(112)은 웨어러블 전자 디바이스(1302)에 통합된다. 멀티-스테이지 타깃 사운드 검출은 웨어러블 전자 디바이스(1302)의 마이크로폰(112)으로부터 수신된 오디오 신호에 기반하여 수행될 수 있다. 웨어러블 전자 디바이스(1302)는 타깃 사운드가 웨어러블 전자 디바이스(1302)에 의해 검출된다는 것을 표시하는 통지를 디스플레이하도록 구성된 디스플레이 스크린, 이를테면 출력 디바이스(250)를 포함한다. 특정 예에서, 출력 디바이스(250)는 타깃 사운드의 검출에 대한 응답으로 햅틱 통지를 제공하는(예컨대, 진동하는) 햅틱 디바이스를 포함한다. 햅틱 통지는 사용자로 하여금 타깃 사운드가 검출되었다는 것을 표시하는 디스플레이된 통지를 보기 위해 웨어러블 전자 디바이스(1302)를 보게 할 수 있다. 따라서, 웨어러블 전자 디바이스(1302)는 청력 손상을 갖는 사용자 또는 헤드셋을 착용한 사용자에게 타깃 사운드가 검출되었음을 경고할 수 있다.

[0067] 도 14는 무선 스피커 및 음성 활성화 디바이스(1400)의 예시적인 예이다. 무선 스피커 및 음성 활성화된 디바이스(1400)는 무선 네트워크 연결성을 가질 수 있고, 어시스턴트 동작을 실행하도록 구성된다. 하나 이상의 프로세서들(160), 마이크로폰(112), 및 하나 이상의 카메라들, 이를테면 카메라(620)는 무선 스피커 및 음성 활성화 디바이스(1400)에 포함된다. 카메라(620)는 통합형 어시스턴트 애플리케이션(1402)에 대한 응답으로, 이를테면, 비디오 컨퍼런스를 개시하라는 사용자 명령에 대한 응답으로, 활성화되도록 구성된다. 카메라(620)는 타깃 사운드 검출기(120) 내의 이진 타깃 사운드 분류기(144)에 의한, 마이크로폰(112)으로부터의 오디오 데이터에서 다수의 타깃 사운드들 중의 임의의 타깃 사운드의 존재의 검출에 응답하여 활성화되도록, 이를테면 타깃 사운드의 검출에 대한 응답으로 감시 카메라로서 기능하도록 추가로 구성된다.

[0068] 무선 스피커 및 음성 활성화된 디바이스(1400)는 또한 스피커(1404)를 포함한다. 동작 동안, 구두 커맨드(verbal command)를 수신하는 것에 대한 응답으로, 무선 스피커 및 음성 활성화 디바이스(1400)는 이를테면, 통합형 어시스턴트 애플리케이션(1402)의 실행을 통해, 어시스턴트 동작들을 실행할 수 있다. 어시스턴트 동작들은, 온도를 조정하는 것, 음악을 플레이하는 것, 조명들을 켜는 것, 비디오 컨퍼런스를 개시하는 것 등을 포함할 수 있다. 예컨대, 어시스턴트 동작들은 키워드(예컨대, "hello assistant") 이후의 커맨드를 수신하는 것에 대한 응답으로 수행된다. 멀티-스테이지 타깃 사운드 검출은 무선 스피커 및 음성 활성화 디바이스(1400)의 마이크로폰(142)으로부터 수신된 오디오 신호에 기반하여 수행될 수 있다. 일부 구현들에서, 통합형 어시스턴트 애플리케이션(1402)은, 타깃 사운드 검출기(120) 내의 이진 타깃 사운드 분류기(144)에 의한, 마이크로폰(112)으로부터의 오디오 데이터에서 다수의 타깃 사운드들 중의 임의의 타깃 사운드의 존재의 검출에 대한 응답으로 활성화된다. 식별된 타깃 사운드(예컨대, 검출기 출력(152))의 표시는 통합형 어시스턴트 애플리케이션(1402)에 제공되고, 통합형 어시스턴트 애플리케이션(1402)은 무선 스피커 및 음성 활성화 디바이스(1400)로 하여금, 통지를 제공하게, 이를테면 스피커(1404)를 통해 가청 스피치 통지를 재생하게, 또는 타깃 사운드(예컨대, 도어의 열림 또는 닫힘)가 무선 스피커 및 음성 활성화 디바이스(1400)에 의해 검출됨을 표시하는 통지를 모바일 디바이스에 송신하게 한다.

[0069] [0082] 도 15를 참조하면, 멀티-스테이지 타깃 사운드 검출의 방법(1500)의 특정 구현이 도시된다. 특정 양상에서, 방법(1500)의 하나 이상의 동작들은, 도 1의 이진 타깃 사운드 분류기(144), 타깃 사운드 검출기(120), 버퍼(130), 프로세서(160), 디바이스(102), 시스템(100), 도 2의 활성화 신호 유닛(204), 다중 타깃 사운드 분류기(210), 활성화 회로부(230), 사운드 콘텍스트 애플리케이션(240), 출력 디바이스(250), 시스템(200), 도 3의 오디오 장면 검출기(302), 오디오 장면 변화 검출기(304), 오디오 장면 분류기(308), 도 4의 장면 전환 분류기(414), 도 5의 계층적 모델 변화 검출기(514), 도 6의 장면 검출기(606) 또는 이들의 조합 중 적어도 하나에 의해 수행된다.

[0070] [0083] 방법(1500)은, 1502에서, 버퍼에 오디오 데이터를 저장하는 단계를 포함한다. 예컨대, 도 1의 버퍼(130)는 도 1을 참조하여 설명한 바와 같이 오디오 데이터(132)를 저장한다. 특정 양상에서, 오디오 데이터(132)는 도 1의 마이크로폰(112)으로부터 수신된 오디오 신호(114)에 대응한다.

[0071] [0084] 방법(1500)은 또한, 1504에서, 타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여 버퍼 내의 오디오 데이터를 프로세싱하는 단계를 포함한다. 예컨대, 도 1의 이진 타깃 사운드 분류기(144)는 도 1을 참조하여 설명된 바와 같이 버퍼(130)에 저장된 오디오 데이터(132)를 프로세싱한다. 이진 타깃 사운드 분류기(144)는 도 1의 타깃 사운드 검출기(150)의 제1 스테이지(140)에 있다.

[0072] [0085] 이 방법(1500)은 1506에서, 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 타깃 사운드 검출

기의 제2 스테이지를 활성화하는 단계를 더 포함한다. 예컨대, 도 1을 참조하여 설명된 바와 같이, 도 1의 제1 스테이지(140)는 제1 스테이지(140)에 의한 타깃 사운드(106)의 검출에 대한 응답으로 타깃 사운드 검출기(120)의 제2 스테이지(150)를 활성화한다. 일부 구현들에서, 이진 타깃 사운드 분류기 및 버퍼는 올웨이즈-온 모드로 동작하고, 제2 스테이지를 활성화하는 단계는, 도 2를 참조하여 설명된 바와 같이, 제1 스테이지로부터 제2 스테이지로 신호를 전송하는 단계, 및 제2 스테이지에서 신호를 수신하는 것에 응답하여 제2 스테이지를 저전력 상태로부터 활성 상태로 전환하는 단계를 포함한다.

[0073] [0086] 방법(1500)은 1508에서, 제2 스테이지의 다중 타깃 사운드 분류기를 사용하여 버퍼로부터의 오디오 데이터를 프로세싱하는 단계를 포함한다. 예컨대, 도 2의 다중 타깃 사운드 분류기(210)는, 도 2를 참조하여 설명된 바와 같이, 제2 스테이지(150)에서 버퍼(130)로부터의 오디오 데이터(132)를 프로세싱한다. 다중 타깃 사운드 분류기는, 예시적인, 비제한적인 예들로서, 사운드 이벤트들의 다수의 클래스들, 이를테면 클래스들(290) 또는 사운드 이벤트 클래스들(812-818)의 세트들 중 하나 이상에 대응하는 다수의 타깃 사운드들에 기반하여 오디오 데이터를 프로세싱할 수 있다.

[0074] [0087] 방법(1500)은 또한, 다수의 타깃 사운드들 각각에 대해, 오디오 데이터에서 그 타깃 사운드의 존재 또는 부재를 표시하는 검출기 출력, 이를테면 검출기 출력(152)을 생성하는 단계를 포함할 수 있다.

[0075] [0088] 일부 구현들에서, 방법(1500)은 또한, 오디오 장면 변화 검출기, 이를테면, 도 3의 오디오 장면 검출기(302)에서 오디오 데이터를 프로세싱하는 단계를 포함한다. 이러한 구현들에서, 오디오 장면 변화를 검출하는 것에 대한 응답으로, 방법(1500)은 오디오 장면 분류기(308)와 같은 오디오 장면 분류기를 활성화하는 단계, 및 오디오 장면 분류기를 사용하여 버퍼로부터의 오디오 데이터를 프로세싱하는 단계를 포함한다. 방법(1500)은 오디오 장면 분류기에서, 클래스들(330)과 같은 다수의 오디오 장면 클래스들에 따라 오디오 데이터를 분류하는 단계를 포함할 수 있다. 예시적인 예에서, 다수의 오디오 장면 클래스들은 집에서, 사무실에서, 식당에서, 자동차에서, 기차에서, 거리에서, 실내에서 또는 실외에서 중 적어도 2개를 포함한다.

[0076] [0089] 오디오 장면 변화를 검출하는 것은, 도 3의 오디오 장면 변화 검출기(304)를 참조하여 설명된 바와 같이, 소음 통계 또는 비-정지형 사운드 통계 중 적어도 하나에서의 변화들을 검출하는 것에 기반할 수 있다. 대안적으로 또는 부가하여, 오디오 장면 변화를 검출하는 것은 장면들 사이의 전환들에 대응하는 오디오 데이터를 사용하여 트레이닝된 분류기, 이를테면 도 4의 장면 전환 분류기(414)를 사용하여 수행될 수 있다. 대안적으로 또는 부가하여, 방법(1500)은 오디오 장면 클래스들의 제1 세트(예컨대, 도 5의 감소된 클래스들의 세트(530))에서 오디오 장면 클래스들 사이의 변화들을 검출하는 것에 기반하여 오디오 장면 변화를 검출하는 단계, 및 오디오 장면 클래스들의 제2 세트(예컨대, 도 3의 클래스들(330))에 따라 오디오 데이터를 분류하는 단계를 포함할 수 있고, 오디오 장면 클래스들의 제1 세트에서의 오디오 장면 클래스들의 제1 카운트(예컨대, 3)는 오디오 장면 클래스들의 제2 세트에서의 오디오 장면 클래스들의 제2 카운트(예컨대, 8)보다 적다.

[0077] [0090] 이진 타깃 사운드 분류기의 프로세싱 동작들은 제2 스테이지 의해 수행되는 프로세싱 동작들과 비교하여 덜 복잡하기 때문에, 이진 타깃 사운드 분류기에서에서 프로세싱되는 오디오 데이터는 제2 스테이지에서 오디오 데이터를 프로세싱하는 것에 비해 더 적은 전력을 소비한다. 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 제2 스테이지를 선택적으로 활성화함으로써, 방법(1500)은 프로세싱 자원들이 보존되고 전체 전력 소비가 감소될 수 있게 한다.

[0078] [0091] 도 15의 방법(1500)은 필드-프로그래밍가능 게이트 어레이(FPGA) 디바이스, 주문형 집적 회로(ASIC), 프로세싱 유닛, 이를테면 중앙 프로세싱 유닛(CPU), DSP, 제어기, 다른 하드웨어 디바이스, 펌웨어 디바이스, 또는 이들의 임의의 조합에 의해 구현될 수 있다. 예로서, 도 15의 방법(1500)은 도 16을 참조하여 설명된 바와 같이 명령들을 실행하는 프로세서에 의해 수행될 수 있다.

[0079] [0092] 도 16을 참조하면, 디바이스의 특정한 예시적 구현의 블록도가 도시되고 포괄적으로 1600으로 지정된다. 다양한 구현들에서, 디바이스(1600)는 도 16에 예시된 것보다 더 많거나 더 적은 컴포넌트들을 가질 수 있다. 예시적인 구현에서, 디바이스(1600)는 디바이스(102)에 대응할 수 있다. 예시적인 구현에서, 디바이스(1600)는 도 1 내지 도 15를 참조하여 설명된 하나 이상의 동작들에 의해 수행될 수 있다.

[0080] [0093] 특정 구현에서, 디바이스(1600)는 프로세서(1606)(예컨대, CPU(central processing unit))를 포함한다. 디바이스(1600)는 하나 이상의 부가적인 프로세서들(1610)(예컨대, 하나 이상의 DSP들)을 포함할 수 있다. 프로세서(1610)는 스피치 및 뮤직 CODEC(coder-decoder)(1608), 타깃 사운드 검출기(120), 사운드 콘텍스트 애플리케이션(240), 활성화 회로부(230), 오디오 장면 검출기(302), 또는 이들의 조합을 포함할 수 있다. 스피치

및 뮤직 코덱(1608)은 음성 코더(“보코더(vocoder)”) 인코더(1636), 보코더 디코더(1638), 또는 이 둘 모두를 포함할 수 있다.

[0081] [0094] 디바이스(1600)는 메모리(1686) 및 CODEC(1634)을 포함할 수 있다. 메모리(1686)는 타깃 사운드 검출기(120), 사운드 콘텍스트 애플리케이션(240), 활성화 회로부(230), 오디오 장면 검출기(302), 또는 이들의 임의의 조합을 참조로 설명된 기능성을 구현하기 위해 하나 이상의 부가적인 프로세서들(1610)(또는 프로세서(1606))에 의해 실행가능한 명령들(1656)을 포함할 수 있다. 메모리(1686)는 버퍼(160)를 포함할 수 있다. 디바이스(1600)는 트랜시버(1650)를 통해 안테나(1652)에 커플링된 무선 제어기(1640)를 포함할 수 있다.

[0082] [0095] 디바이스(1600)는 디스플레이 제어기(1626)에 커플링된 디스플레이(1628)를 포함할 수 있다. 스피커(1692) 및 마이크로폰(112)이 CODEC(1634)에 커플링될 수 있다. CODEC(1634)은 디지털-아날로그 변환기(1602) 및 아날로그-디지털 변환기(1604)를 포함할 수 있다. 특정 구현에서, CODEC(1634)은 마이크로폰(112)으로부터 아날로그 신호들을 수신하고, 아날로그-디지털 변환기(1604)를 사용하여 아날로그 신호들을 디지털 신호들로 변환하고, 디지털 신호들을 스피치 및 뮤직 코덱(1608)에 제공할 수 있다. 스피치 및 뮤직 코덱(1608)은 디지털 신호들을 프로세싱할 수 있고, 디지털 신호들은 타깃 사운드 검출기(120) 및 오디오 장면 검출기(302) 중 하나 이상에 의해 추가로 프로세싱될 수 있다. 특정 구현에서, 스피치 및 뮤직 CODEC(1608)은 CODEC(1634)에 디지털 신호들을 제공할 수 있다. CODEC(1634)은 디지털-아날로그 변환기(1602)를 사용하여 디지털 신호들을 아날로그 신호들로 변환할 수 있고, 아날로그 신호들을 스피커(1692)에 제공할 수 있다.

[0083] [0096] 특정 구현에서, 디바이스(1600)는 시스템-인-패키지 또는 시스템-온-칩 디바이스(1622)에 포함될 수 있다. 특정 구현에서, 메모리(1686), 프로세서(1606), 프로세서들(1610), 디스플레이 제어기(1626), CODEC(1634), 및 무선 제어기(1640)는 시스템-인-패키지 또는 시스템-온-칩 디바이스(1622)에 포함된다. 특정한 구현에서, 입력 디바이스(1630) 및 전력 공급부(1644)는, 시스템-온-칩 디바이스(1622)에 커플링된다. 또한, 특정한 구현에서, 도 16에 예시된 바와 같이, 디스플레이(1628), 입력 디바이스(1630), 스피커(1692), 마이크로폰(112), 안테나(1652), 및 전력 공급부(1644)는 시스템-온-칩 디바이스(1622) 외부에 있다. 특정 구현에서, 디스플레이(1628), 입력 디바이스(1630), 스피커(1692), 마이크로폰(112), 안테나(1652), 및 전력 공급부(1644) 각각은, 인터페이스 또는 제어기와 같은 시스템-온-칩 디바이스(1622)의 컴포넌트에 커플링될 수 있다.

[0084] [0097] 디바이스(1600)는 스마트 스피커, 스피커 바, 모바일 통신 디바이스, 스마트 폰, 셀룰러 폰, 랩톱 컴퓨터, 컴퓨터, 태블릿 컴퓨터, 개인 휴대 정보 단말, 디스플레이 디바이스, 텔레비전, 게임 콘솔, 뮤직 플레이어, 라디오, 디지털 비디오 플레이어, DVD(digital video disc) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 운송수단, 헤드셋, 증강 현실 헤드셋, 가상 현실 헤드셋, 비행체, 또는 이들의 임의의 조합을 포함할 수 있다.

[0085] [0098] 설명된 구현들과 함께, 입력 사운드를 나타내는 오디오 신호를 프로세싱하기 위한 장치는 타깃 사운드를 검출하기 위한 수단을 포함한다. 타깃 사운드를 검출하기 위한 수단은 제1 스테이지 및 제2 스테이지를 포함한다. 제1 스테이지는, 오디오 데이터의 이진 타깃 사운드 분류를 생성하고, 그리고 타깃 사운드를 포함하는 것으로서의 오디오 데이터의 분류에 대한 응답으로 제2 스테이지를 활성화하기 위한 수단을 포함한다. 예컨대, 타깃 사운드를 검출하기 위한 수단은 타깃 사운드 검출기(120), 하나 이상의 프로세서들(160), 하나 이상의 프로세서들(1610), 타깃 사운드를 검출하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응한다. 이진 타깃 사운드 분류를 생성하고 제2 스테이지를 활성화하기 위한 수단은, 이진 타깃 사운드 분류기(144), 이진 타깃 사운드 분류를 생성하고 제2 스테이지를 활성화하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.

[0086] [0099] 장치는 또한, 오디오 데이터를 버퍼링하고, 그리고 타깃 사운드를 포함하는 것으로서의 오디오 데이터의 분류에 대한 응답으로 오디오 데이터를 제2 스테이지에 제공하기 위한 수단을 포함한다. 예컨대, 오디오 데이터를 버퍼링하고 오디오 데이터를 제2 스테이지에 제공하기 위한 수단은, 버퍼(160), 하나 이상의 프로세서들(160), 하나 이상의 프로세서들(1610), 오디오 데이터를 버퍼링하도록 구성되고 타깃 사운드를 포함하는 것으로서의 오디오 데이터의 분류에 대한 응답으로 제2 스테이지에 오디오 데이터를 제공하는 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.

[0087] [0100] 일부 구현들에서, 장치는 오디오 장면을 검출하기 위한 수단을 더 포함하고, 오디오 장면을 검출하기 위한 수단은 오디오 데이터에서 오디오 장면 변화를 검출하기 위한 수단, 및 오디오 장면 변화의 검출에 대한 응답으로 오디오 데이터를 특정 오디오 장면으로서 분류하기 위한 수단을 포함한다. 예컨대, 오디오 장면을 검출하기 위한 수단은 오디오 장면 검출기(302), 하나 이상의 프로세서들(160), 하나 이상의 프로세서들(1610), 오디오 장면을 검출하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할

수 있다. 오디오 데이터에서 오디오 장면 변화를 검출하기 위한 수단은 오디오 장면 변화 검출기(304), 장면 전환 분류기(414), 계층적 모델 변화 검출기(514), 오디오 데이터에서 오디오 장면 변화를 검출하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다. 오디오 장면 변화의 검출에 대한 응답으로 오디오 데이터를 특정한 오디오 장면으로서 분류하기 위한 수단은 오디오 장면 분류기(308), 오디오 장면 변화의 검출에 대한 응답으로 오디오 데이터를 특정한 오디오 장면으로서 분류하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 이들의 임의의 조합에 대응할 수 있다.

[0088] [0101] 일부 구현들에서, 비-일시적인 컴퓨터-판독가능 매체(예컨대, 메모리(1686))는 명령들(예컨대, 명령들(1656))을 포함하며, 명령들은 하나 이상의 프로세서들(예컨대, 하나 이상의 프로세서들(1610) 또는 프로세서(1606))에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 오디오 데이터를 버퍼(예컨대, 버퍼(130))에 저장하고, 그리고 타깃 사운드 검출기의 제1 스테이지(예컨대, 타깃 사운드 검출기(120)의 제1 스테이지(140))에서 이진 타깃 사운드 분류기(예컨대, 이진 타깃 사운드 분류기(144))를 사용하여 버퍼 내의 오디오 데이터를 프로세싱하기 위한 동작들을 수행하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때 또한, 하나 이상의 프로세서들로 하여금, 제1 스테이지에 의한 타깃 사운드의 검출에 대한 응답으로 타깃 사운드 검출기의 제2 스테이지(예컨대, 제2 스테이지(150))를 활성화하게 하고, 그리고 제2 스테이지에서 다중 타깃 사운드 분류기(예컨대, 다중 타깃 사운드 분류기(210))를 사용하여 버퍼로부터의 오디오 데이터를 프로세싱하게 한다.

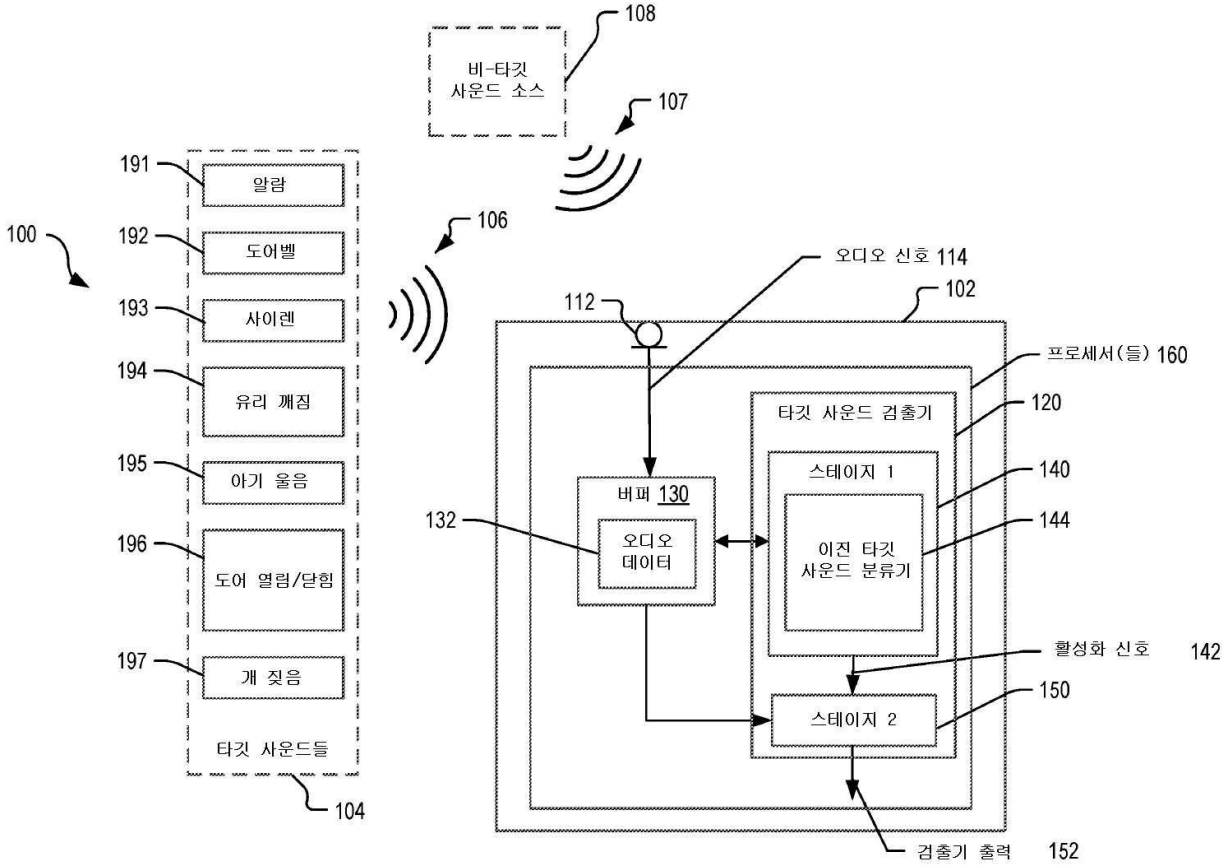
[0089] [0102] 당업자들은, 본 명세서에 개시된 구현들과 관련하여 설명된 다양한 예시적인 로직 블록들, 구성들, 모듈들, 회로들, 및 알고리즘 단계들이 전자 하드웨어, 프로세서에 의해 실행되는 컴퓨터 소프트웨어, 또는 둘 모두의 결합들로서 구현될 수 있음을 추가적으로 인식할 것이다. 다양한 예시적인 컴포넌트들, 블록들, 구성들, 모듈들, 회로들, 및 단계들은 이들의 기능 관점들에서 일반적으로 상술되었다. 이러한 기능이 하드웨어로 구현되는지 또는 프로세서 실행가능 명령들로 구현되는지 여부는 특정 애플리케이션, 및 전체 시스템에 부과된 설계 제약들에 의존한다. 당업자들은 설명된 기능을 각각의 특정 애플리케이션에 대해 다양한 방식들로 구현할 수 있고, 이러한 구현 결정들이 본 개시의 범주를 벗어나는 것으로 해석되어서는 안 된다.

[0090] [0103] 본 명세서에 기재된 구현들과 관련하여 설명된 알고리즘 또는 방법의 단계들은 직접 하드웨어로, 프로세서에 의해 실행되는 소프트웨어 모듈로, 또는 이 둘의 결합으로 구현될 수도 있다. 소프트웨어 모듈은, 랜덤 액세스 메모리(RAM), 플래시 메모리, 판독-전용 메모리(ROM), 프로그래밍가능 판독-전용 메모리(PROM), 소거가능한 프로그래밍가능 판독-전용 메모리(EPROM), 전기적으로 소거가능한 프로그래밍가능 판독-전용 메모리(EEPROM), 레지스터들, 하드 디스크, 착탈형 디스크, 컴팩트 디스크 판독-전용 메모리(CD-ROM), 또는 당업계에 알려진 임의의 다른 형태의 비일시적 저장 매체에 상주할 수 있다. 예시적인 저장 매체는, 프로세서가 저장 매체로부터 정보를 판독하고 저장 매체에 정보를 기입할 수 있도록 프로세서에 커플링된다. 대안적으로, 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.

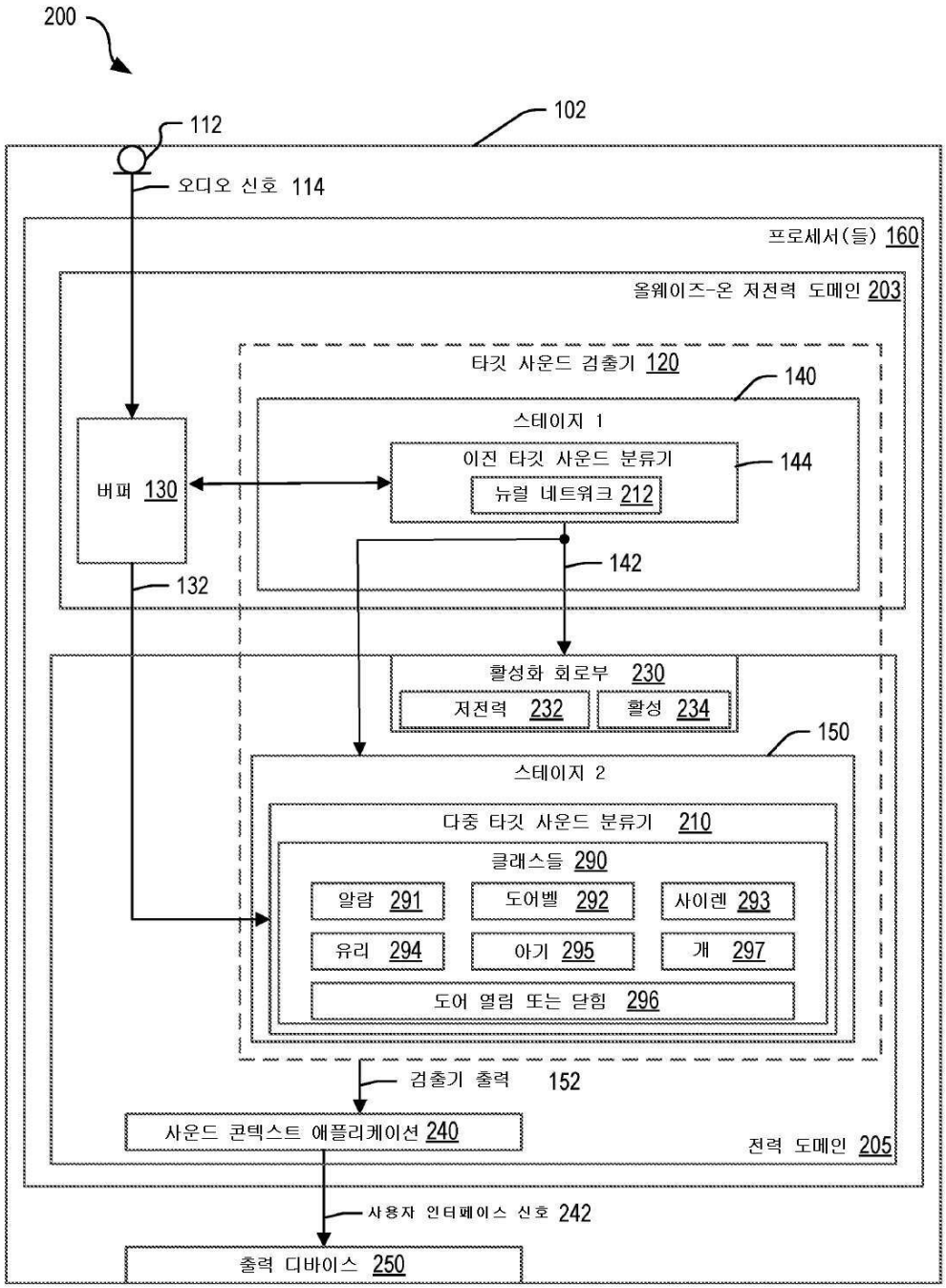
[0091] [0104] 개시된 구현들의 이전 설명은 당업자가 개시된 구현들을 실시하거나 또는 사용할 수 있도록 제공된다. 이러한 구현들에 대한 다양한 변형들은 당업자들에게 용이하게 명백할 것이며, 본 명세서에 정의된 원리들은 본 개시의 범위를 벗어나지 않으면서 다른 구현들에 적용될 수 있다. 따라서, 본 개시내용은 본 명세서에 나타낸 구현들로 제한되도록 의도되지 않으며, 다음의 청구항들에 의해 정의된 바와 같은 원리들 및 신규한 특성들과 일치하는 가능한 가장 넓은 범위에 부합할 것이다.

## 도면

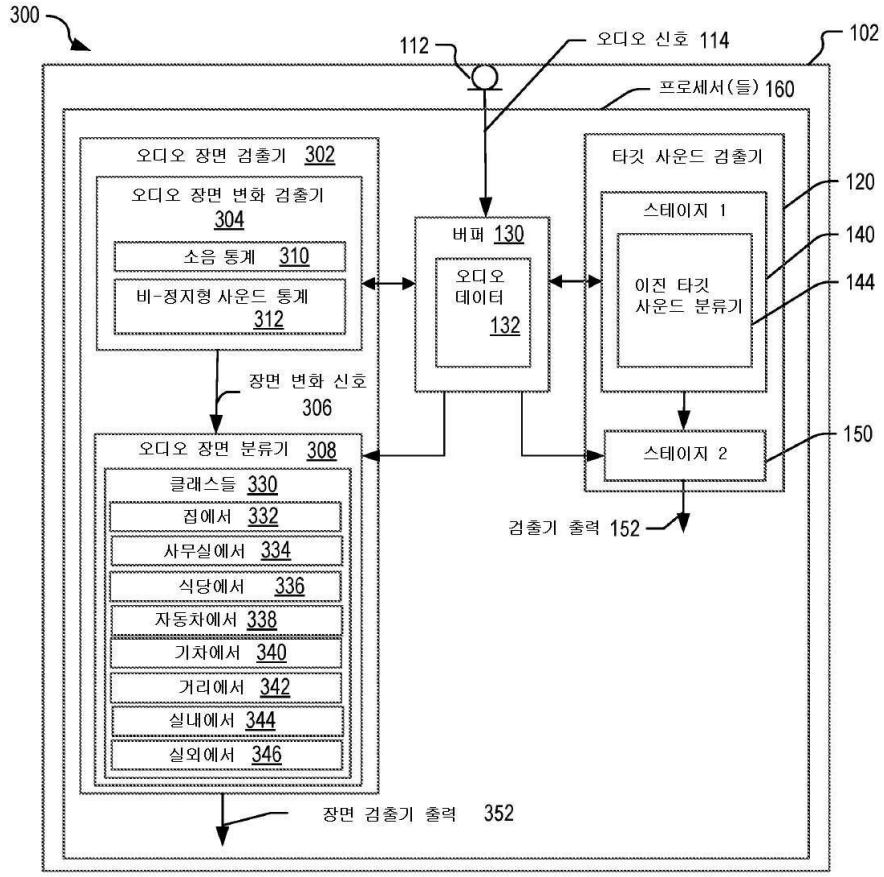
### 도면1


비-타깃
사운드 소스
108
107
106
100
191
알람
192
도어벨
193
사이렌
194
유리 깨짐
195
아기 울음
196
도어 열림/닫힘
197
개 짖음
타깃 사운드들
104
112
오디오 신호 114
102
프로세서(들) 160
타깃 사운드 검출기
120
스테이지 1
140
버퍼 130
132
오디오
데이터
이진 타깃
사운드 분류기
144
활성화 신호
142
스테이지 2
150
검출기 출력
152

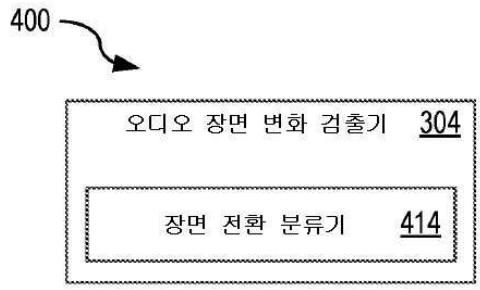
도면2


200
112
102
오디오 신호 114
프로세서(들) 160
올웨이즈-온 저전력 도메인 203
타깃 사운드 검출기 120
140
스테이지 1
이진 타깃 사운드 분류기
144
뉴럴 네트워크 212
버퍼 130
142
132
활성화 회로부 230
저전력 232
활성 234
150
스테이지 2
다중 타깃 사운드 분류기 210
클래스들 290
알람 291
도어벨 292
사이렌 293
유리 294
아기 295
개 297
도어 열림 또는 닫힘 296
검출기 출력 152
사운드 콘텍스트 애플리케이션 240
전력 도메인 205
사용자 인터페이스 신호 242
출력 디바이스 250

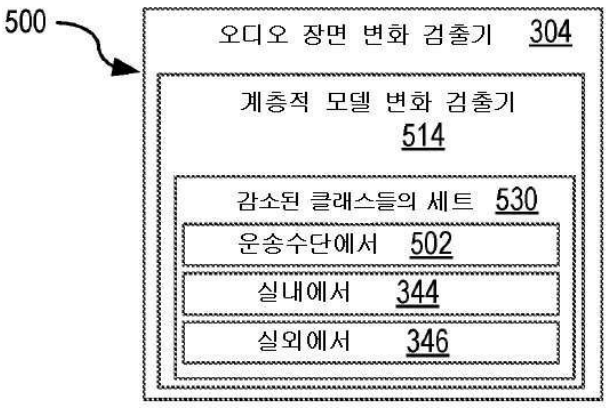
도면3


300
112
오디오 신호 114
102
프로세서(들) 160
오디오 장면 검출기 302
오디오 장면 변화 검출기 304
소음 통계 310
비-정지형 사운드 통계 312
버퍼 130
오디오 데이터 132
타깃 사운드 검출기
스테이지 1
이진 타깃 사운드 분류기
120
140
144
장면 변화 신호 306
오디오 장면 분류기 308
클래스들 330
집에서 332
사무실에서 334
식당에서 336
자동차에서 338
기차에서 340
거리에서 342
실내에서 344
실외에서 346
스테이지 2
150
검출기 출력 152
장면 검출기 출력 352

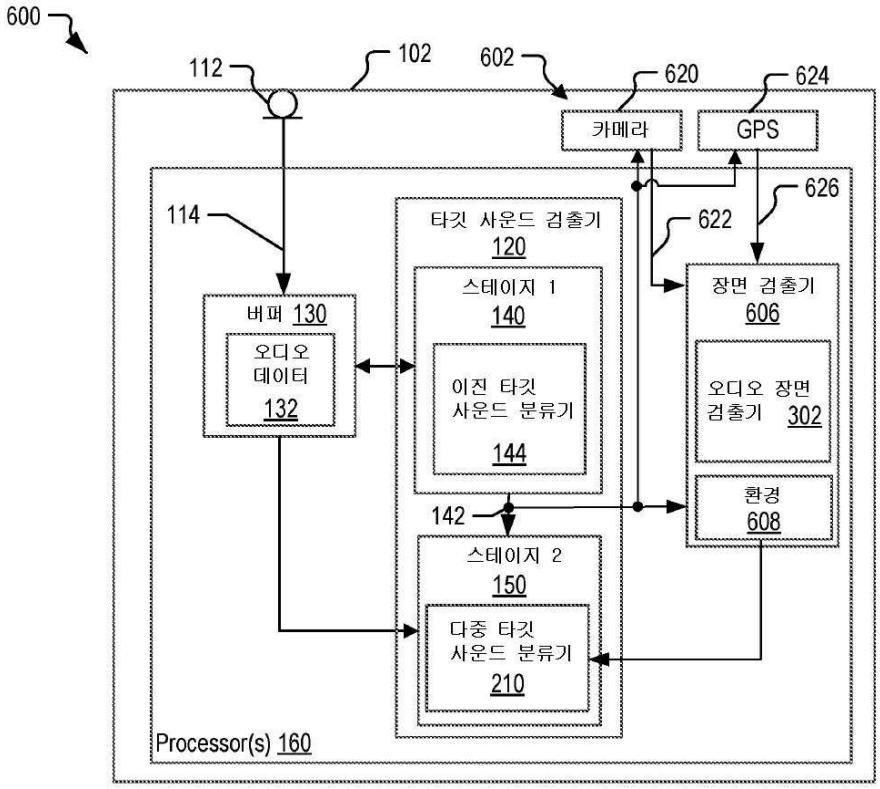
도면4


400
오디오 장면 변화 검출기 304
장면 전환 분류기 414

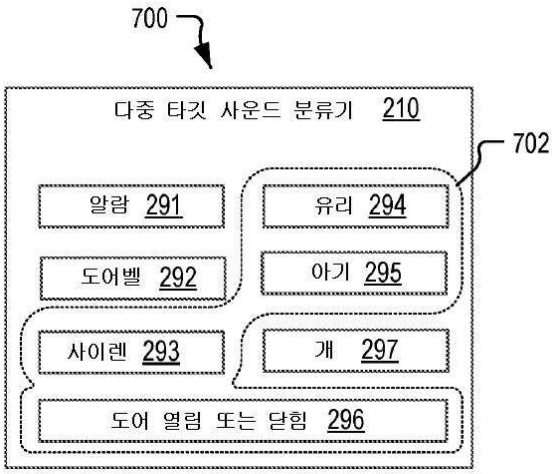
도면5


500
오디오 장면 변화 검출기 304
계층적 모델 변화 검출기 514
감소된 클래스들의 세트 530
운송수단에서 502
실내에서 344
실외에서 346

도면6


600
112
102
602
620
624
카메라
GPS
626
114
622
타깃 사운드 검출기
120
스테이지 1
140
장면 검출기
606
버퍼 130
오디오
데이터
132
이진 타깃
사운드 분류기
144
오디오 장면
검출기 302
환경
608
142
스테이지 2
150
다중 타깃
사운드 분류기
210
Processor(s) 160

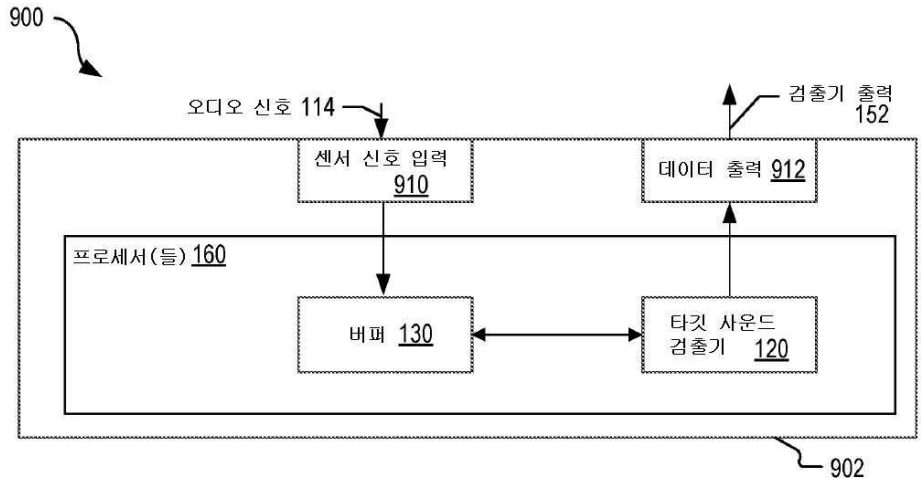
도면7


700
다중 타깃 사운드 분류기 210
702
알람 291
유리 294
도어벨 292
아기 295
사이렌 293
개 297
도어 열림 또는 닫힘 296

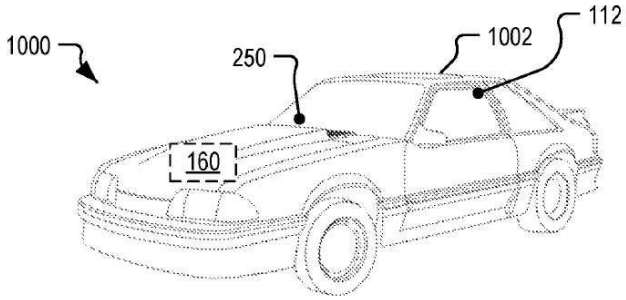
도면8


800
다중 타깃 사운드 분류기 210
트레이닝된 데이터의 세트 802
사운드 이벤트 클래스들의 세트 812
트레이닝된 데이터의 세트 804
사운드 이벤트 클래스들의 세트 814
트레이닝된 데이터의 세트 808
사운드 이벤트 클래스들의 세트 818

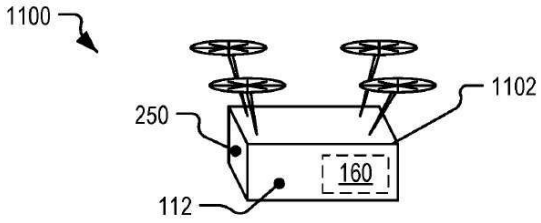
도면9


900
오디오 신호 114
검출기 출력 152
센서 신호 입력 910
데이터 출력 912
프로세서(들) 160
버퍼 130
타깃 사운드 검출기 120
902

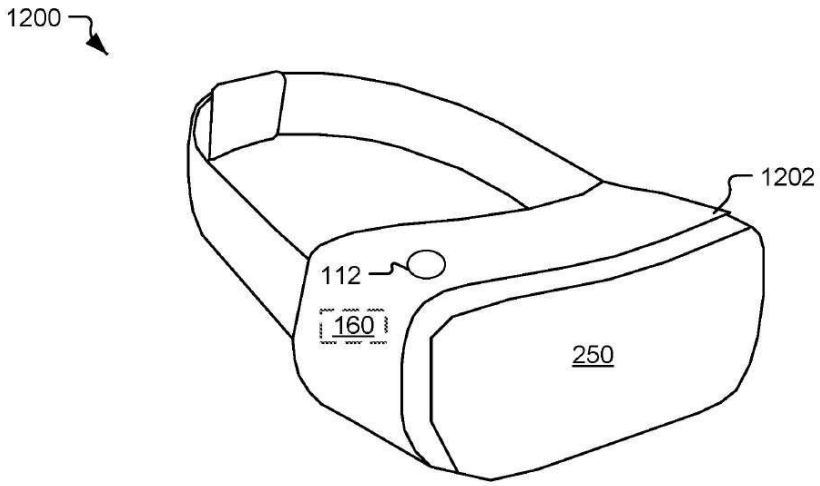
도면10


1000
250
1002
112
160

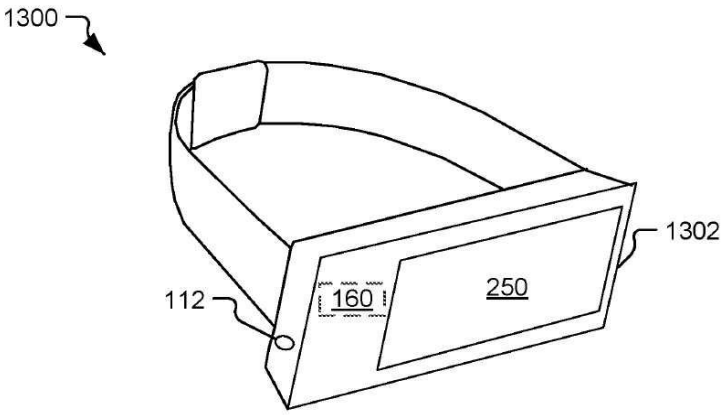
도면11



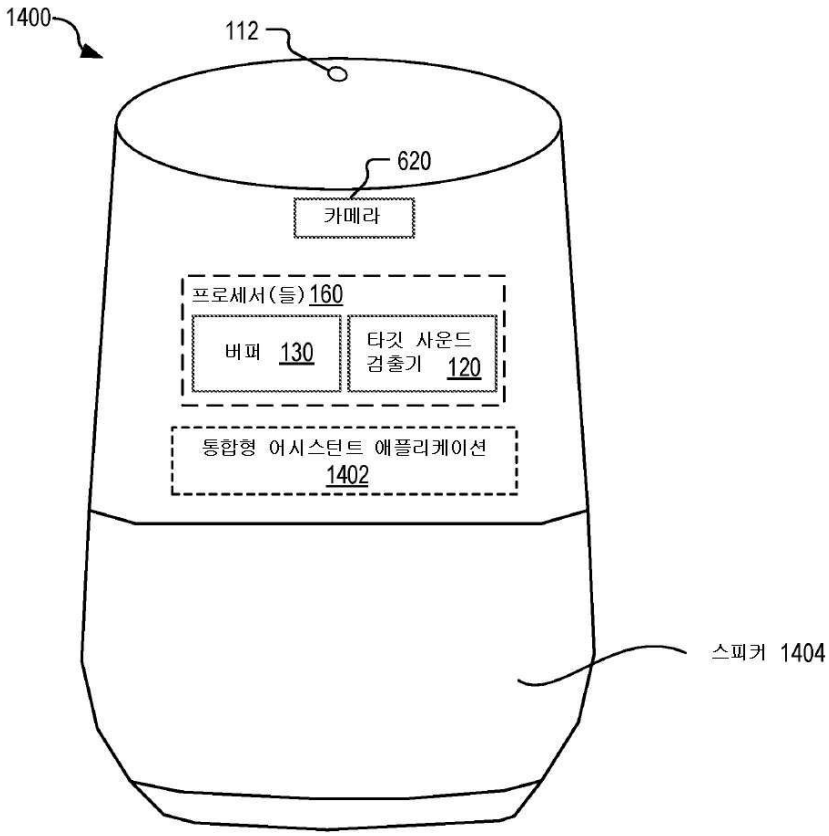

도면12



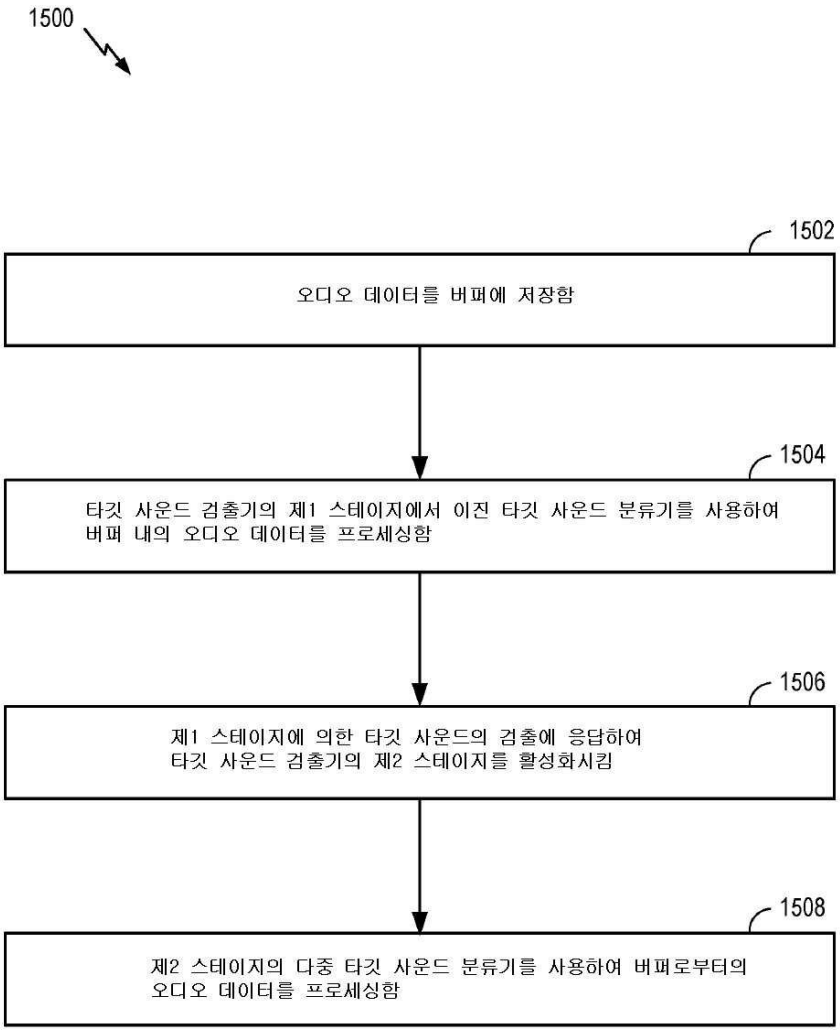

도면13



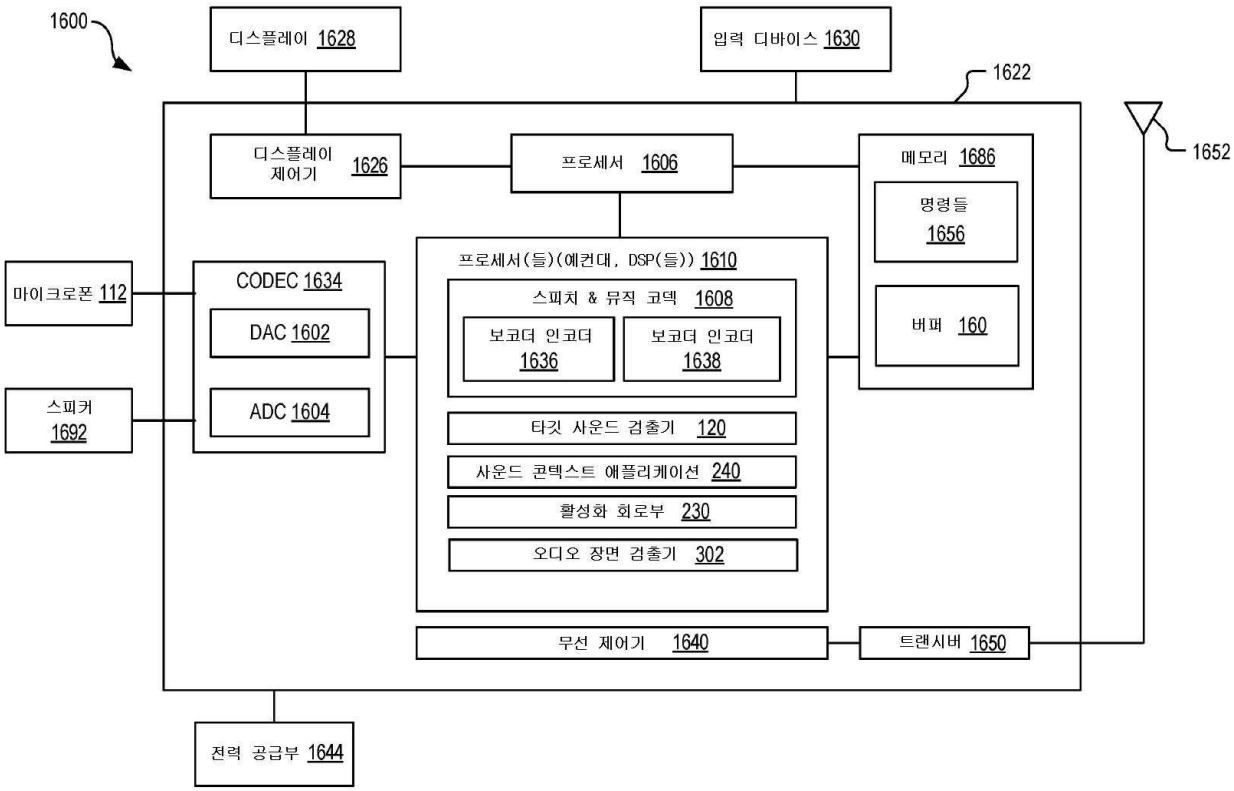
도면14


1400
112
620
카메라
프로세서(들) 160
버퍼 130
타깃 사운드 검출기 120
통합형 어시스턴트 애플리케이션 1402
스피커 1404

도면15


1500
1502
오디오 데이터를 버퍼에 저장함
1504
타깃 사운드 검출기의 제1 스테이지에서 이진 타깃 사운드 분류기를 사용하여
버퍼 내의 오디오 데이터를 프로세싱함
1506
제1 스테이지에 의한 타깃 사운드의 검출에 응답하여
타깃 사운드 검출기의 제2 스테이지를 활성화시킴
1508
제2 스테이지의 다중 타깃 사운드 분류기를 사용하여 버퍼로부터의
오디오 데이터를 프로세싱함

도면16


1600
디스플레이 1628
입력 디바이스 1630
1622
1652
디스플레이 제어기 1626
프로세서 1606
메모리 1686
명령들 1656
버퍼 160
프로세서(들)(예컨대, DSP(들)) 1610
스피치 & 뮤직 코덱 1608
보코더 인코더 1636
보코더 인코더 1638
마이크로폰 112
CODEC 1634
DAC 1602
ADC 1604
스피커 1692
타깃 사운드 검출기 120
사운드 콘텍스트 애플리케이션 240
활성화 회로부 230
오디오 장면 검출기 302
무선 제어기 1640
트랜시버 1650
전력 공급부 1644