(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0084154
(43) 공개일자 2023년06월12일

(51) 국제특허분류(Int. Cl.)
G10L 25/78 (2013.01) *G06F 1/3206* (2019.01)
G06F 1/3287 (2019.01) *G10L 15/12* (2006.01)
G10L 15/22 (2006.01) *H04R 3/00* (2006.01)
(52) CPC특허분류
G10L 25/78 (2013.01)
G06F 1/3206 (2019.01)
(21) 출원번호 10-2023-7011211
(22) 출원일자(국제) 2021년09월17일
심사청구일자 없음
(85) 번역문제출일자 2023년03월31일
(86) 국제출원번호 PCT/US2021/071503
(87) 국제공개번호 WO 2022/076963
국제공개일자 2022년04월14일
(30) 우선권주장
63/089,507 2020년10월08일 미국(US)
17/308,593 2021년05월05일 미국(US)

(71) 출원인
퀄컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌 디에고 모어하우스 드라이브 5775
(72) 발명자
샤바지 미르자하산루 타헤르
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775
알베스 호제리우 게데스
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775
(뒷면에 계속)
(74) 대리인
특허법인코리아나

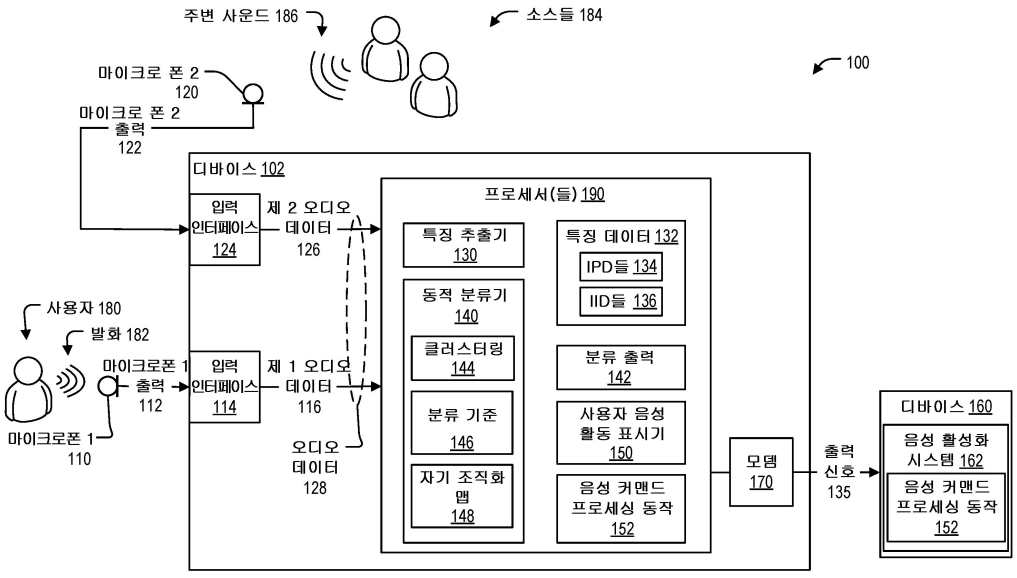
전체 청구항 수 : 총 30 항

(54) 발명의 명칭 **동적 분류기를 사용한 사용자 음성 활동 검출**

(57) 요 약

디바이스는 명령들을 저장하도록 구성된 메모리 및 명령들을 실행하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 명령을 수행하도록 구성된다. 하나 이상의 프로세서들은 또한 오디오 데이터를 동적 분류기에 제공하는 명령들을 실행하도록 구성된다. 동적 분류기는 오디오 데이터에 대응하는 분류 출력을 생성하도록 구성된다. 하나 이상의 프로세서들은 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 명령들을 실행하도록 추가로 구성된다.

대 표 도



(52) CPC특허분류
G06F 1/3287 (2019.01)
G10L 15/12 (2013.01)
G10L 15/22 (2013.01)
H04R 3/005 (2013.01)
Y02D 10/00 (2020.08)
(72) 발명자
피써 에릭
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775
김 래훈
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775

명 세 서

청구범위

청구항 1

디바이스로서,

명령들을 저장하도록 구성된 메모리; 및

하나 이상의 프로세서들을 포함하고,

상기 하나 이상의 프로세서들은:

제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고;

상기 오디오 데이터에 대응하는 분류 출력을 생성하도록 구성된 동적 분류기에 상기 오디오 데이터를 제공하고;

상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 명령들을 실행하도록 구성되는, 디바이스.

청구항 2

제 1 항에 있어서,

상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 사용자의 발화들을 캡처하도록 구성되고, 상기 제 2 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 주변 사운드를 캡처하도록 구성되는, 디바이스.

청구항 3

제 1 항에 있어서,

상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초하는, 디바이스.

청구항 4

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하도록 추가로 구성되고, 상기 오디오 데이터는 상기 특징 데이터로서 상기 동적 분류기에 제공되고, 상기 분류 출력은 상기 특징 데이터에 기초하는, 디바이스.

청구항 5

제 4 항에 있어서,

상기 특징 데이터는,

상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및

상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함하는, 디바이스.

청구항 6

제 4 항에 있어서,

상기 하나 이상의 프로세서들은 상기 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초하여 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성되는, 디바이스.

청구항 7

제 4 항에 있어서,

상기 하나 이상의 프로세서들은 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하도록 추가로 구성되고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함하는, 디바이스.

청구항 8

제 4 항에 있어서,

상기 동적 분류기는 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 적응적으로 클러스터링하도록 구성되는, 디바이스.

청구항 9

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 상기 오디오 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하도록 추가로 구성되는, 디바이스.

청구항 10

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 상기 동적 분류기의 분류 결정 기준을 업데이트하도록 추가로 구성되는, 디바이스.

청구항 11

제 1 항에 있어서,

상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함하는, 디바이스.

청구항 12

제 1 항에 있어서,

상기 동적 분류기는 오디오 데이터의 세트들의 시퀀스를 수신하고, 상기 시퀀스에서의 오디오 데이터의 이전 세트에 적어도 부분적으로 기초하여 상기 시퀀스의 세트를 적응적으로 클러스터링하도록 추가로 구성되는, 디바이스.

청구항 13

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하도록 추가로 구성되는, 디바이스.

청구항 14

제 13 항에 있어서,

상기 하나 이상의 프로세서들은, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하도록 구성되는, 디바이스.

청구항 15

제 14 항에 있어서,

상기 하나 이상의 프로세서들은, 추가로,

상기 동적 분류기를 포함하는 올웨이즈온 (always-on) 전력 도메인; 및

음성 커맨드 프로세싱 유닛을 포함하는 제 2 전력 도메인을 포함하고,

상기 웨이크업 신호는 상기 음성 커맨드 프로세싱 유닛을 활성화하기 위해 저전력 모드로부터 상기 제 2 전력 도메인을 전이하도록 구성되는, 디바이스.

청구항 16

제 1 항에 있어서,

상기 하나 이상의 프로세서들에 커플링된 모뎀을 추가로 포함하고, 상기 모뎀은, 상기 오디오 데이터가 상기 동적 분류기에 기초한 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하도록 구성되는, 디바이스.

청구항 17

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2 마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기 위해 사용자의 입에 상기 제 2 마이크로폰보다 가깝게 상기 제 1 마이크로폰을 위치설정하도록 구성되는, 디바이스.

청구항 18

제 1 항에 있어서,

하나 이상의 프로세서들은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 또는 증강 현실 헤드셋 중 적어도 하나에 통합되는, 디바이스.

청구항 19

제 1 항에 있어서,

상기 하나 이상의 프로세서들은 비히클에 통합되고, 상기 비히클은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 비히클의 조작자의 발화들을 캡처하도록 위치설정되는, 디바이스.

청구항 20

음성 활동 검출 방법으로서,

하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계;

하나 이상의 프로세서들에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 상기 오디오 데이터를 동적 분류기에 제공하는 단계; 및

하나 이상의 프로세서들에서 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함하는, 음성 활동 검출 방법.

청구항 21

제 20 항에 있어서,

상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초하는, 음성 활동 검출 방법.

청구항 22

제 20 항에 있어서,

상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함하는, 음성 활동 검출 방법.

청구항 23

제 20 항에 있어서,

상기 오디오 데이터가 상기 사용자 음성 활동에 대응하는지 여부를 결정하는 단계는 상기 오디오 데이터에 대응하는 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초하는, 음성 활동 검출 방법.

청구항 24

제 20 항에 있어서,

상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하는 단계를 추가로 포함하는, 음성 활동 검출 방법.

청구항 25

제 24 항에 있어서,

상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하는 단계를 추가로 포함하는, 음성 활동 검출 방법.

청구항 26

제 20 항에 있어서,

상기 오디오 데이터가 상기 동적 분류기에 기초하여 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하는 단계를 추가로 포함하는, 음성 활동 검출 방법.

청구항 27

명령들을 포함하는 비일시적 컴퓨터 판독가능 저장 매체로서,

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

제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고;

상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 동적 분류기에 상기 오디오 데이터를 제공하고;

상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하게 하는, 비일시적 컴퓨터 판독가능 저장 매체.

청구항 28

제 27 항에 있어서,

상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초하는, 비일시적 컴퓨터 판독가능 저장 매체.

청구항 29

장치로서,

제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하기 위한 수단;

동적 분류기에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위한 수단; 및

상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를

결정하기 위한 수단을 포함하는, 장치.

청구항 30

제 29 항에 있어서,

상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초하는, 장치.

발명의 설명

기 술 분 야

[0001] ***I. 관련 출원들에 대한 상호참조***

[0002] 본 출원은 2020년 10월 8일자로 출원된 공동 소유 미국 가특허 출원 제63/089,507호, 및 2021년 5월 5일자로 출원된 미국 정규 특허 출원 제 17/308,593호로부터의 우선권의 이익을 주장하며, 이들 각각의 내용들은 본 명세서에 참조로 전부 명백히 통합된다.

[0003] ***II. 분야***

[0004] 본 개시는 일반적으로 자기 음성 활동 검출에 관한 것이다.

배 경 기 술

[0005] **III. 관련 기술의 설명**

[0006] 기술의 진보는 더 소형이고 더 강력한 컴퓨팅 디바이스들을 발생시켰다. 예를 들어, 소형이고 경량이며 사용자들에 의해 용이하게 휴대되는 모바일 및 스마트 폰들과 같은 무선 전화기들, 태블릿들 및 랩톱 컴퓨터들을 포함하는 다양한 휴대용 개인 컴퓨팅 디바이스들이 현재 존재한다. 이들 디바이스들은 무선 네트워크들을 통해 보이스 및 데이터 패킷들을 통신할 수 있다. 추가로, 다수의 그러한 디바이스들은 디지털 스틸 카메라, 디지털 비디오 카메라, 디지털 레코더, 및 오디오 파일 플레이어와 같은 추가적인 기능성을 통합한다. 또한, 그러한 디바이스들은, 인터넷에 액세스하는데 사용될 수 있는 웹 브라우저 애플리케이션과 같은 소프트웨어 애플리케이션들을 포함한 실행가능 명령들을 프로세싱할 수 있다. 그에 따라, 이들 디바이스들은 상당한 컴퓨팅 능력들을 포함할 수 있다.

[0007] 이러한 컴퓨팅 디바이스들은 종종 하나 이상의 마이크로폰들로부터 오디오 신호를 수신하는 기능을 통합한다. 예를 들어, 오디오 신호는 마이크로폰들에 의해 캡처된 사용자 스피치, 마이크로폰들에 의해 캡처된 외부 사운드들, 또는 그 조합을 나타낼 수 있다. 예시를 위해, 헤드셋 디바이스는 사용자의 스피치 (예를 들어, 헤드셋을 착용한 사람이 말하는 스피치) 와 다른 소스들로부터 기원되는 스피치를 구별하기 위한 노력으로 자기 음성 활동 검출을 포함할 수 있다. 예를 들어, 헤드셋 디바이스를 포함하는 시스템이 키워드 활성화를 지원할 때, 자기-음성 활동 검출은 근처의 사람들로부터 기원되는 스피치 ("비-사용자 스피치" 로 지칭됨) 에 기초하여 하나 이상의 컴포넌트들 또는 동작들의 활성화가 개시되는 "오경보들" 을 감소시킬 수 있다. 이러한 오경보를 감소시키는 것은 디바이스의 전력 소비 효율을 향상시킨다. 그러나, 사용자 스피치와 비-사용자 스피치를 구별하기 위해 오디오 신호 프로세싱을 수행하는 것은 또한 전력을 소비하고, 사용자 스피치와 비-사용자 스피치를 구별함에 있어서 디바이스의 정확도를 개선하기 위한 종래의 기술들은 또한 디바이스의 전력 소비 및 프로세싱 리소스 요건들을 증가시키는 경향이 있다.

발명의 내용

[0008] ***IV. Summary***

[0009] 본 개시의 하나의 구현예에 따르면, 디바이스는 명령들을 저장하도록 구성된 메모리 및 명령들을 실행하도록 구성된 하나 이상의 프로세서들을 포함한다. 하나 이상의 프로세서들은, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 명령을 수행하도록 구성된다. 하나 이상의 프로세서들은 또한 오디오 데이터를 동적 분류기에 제공하는 명령들을 실행하도록 구성된다. 동적 분류기는 오디오 데이터에 대응하는 분류 출력을 생성하도록 구성된다. 하나 이상의 프로세서들은 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가

사용자 음성 활동에 대응하는지 여부를 결정하는 명령들을 실행하도록 추가로 구성된다.

[0010] 본 개시의 또 다른 구현예에 따르면, 방법은 하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계를 포함한다. 방법은 추가로 하나 이상의 프로세서들에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 상기 오디오 데이터를 동적 분류기에 제공하는 단계를 포함한다. 방법은 또한 하나 이상의 프로세서들에서 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함한다.

[0011] 본 개시의 또 다른 구현에 따르면, 비일시적 컴퓨터 판독가능 저장 매체는, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하게 하는 명령들을 포함한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 추가로, 하나 이상의 프로세서들로 하여금, 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 오디오 데이터를 동적 분류기에 제공하게 한다.

명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 또한, 하나 이상의 프로세서들로 하여금, 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하게 한다.

[0012] 본 개시의 또 다른 구현예에 따르면, 장치는 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 수신하기 위한 수단을 포함한다. 장치는 추가로 동적 분류기에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위한 수단을 포함한다. 장치는 또한 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단을 포함한다.

[0013] 본 발명의 다른 양태들, 이점들 및 특징들은 다음 섹션들, 즉 도면의 간단한 설명, 상세한 설명 및 청구항들을 포함하여, 본원 전체를 검토한 후 분명해질 것이다.

도면의 간단한 설명

[0014] ***V. 도면의 간단한 설명***

도 1 은 본 개시의 일부 예들에 따른, 자기-음성 활동 검출을 수행하도록 동작가능한 시스템의 특정한 예시적인 양태의 블록도이다.

도 2 는 본 개시의 일부 예들에 따른, 자기 음성 활동 검출과 연관된 동작들의 예시적인 양태의 다이어그램이다.

도 3 은 본 개시의 일부 예들에 따른, 자기 음성 활동 검출을 수행하도록 동작가능한 시스템의 예시적인 양태의 블록도이다.

도 4 는 본 개시의 일부 예들에 따른, 도 1 의 시스템의 컴포넌트들의 동작의 예시적인 양태의 다이어그램이다.

도 5 는 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 집적 회로의 예를 예시한다.

도 6 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 모바일 디바이스의 다이어그램이다.

도 7 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 헤드셋의 다이어그램이다.

도 8 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 웨어러블 전자 디바이스의 다이어그램이다.

도 9 는 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 음성 제어 스피커 시스템의 다이어그램이다.

도 10 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 카메라의 다이어그램이다.

도 11 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 가상 현실

또는 증강 현실 헤드셋과 같은 헤드셋의 다이어그램이다.

도 12 는 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 비히클의 제 1 예의 다이어그램이다.

도 13 은 본 개시의 일부 예들에 따른, 사용자 음성 활동을 검출하기 위한 동적 분류기를 포함하는 비히클의 제 2 예의 다이어그램이다.

도 14a 는 본 개시의 일부 예들에 따른, 도 1 의 디바이스에 의해 수행될 수도 있는 자기 음성 활동 검출의 방법의 특정 구현의 다이어그램이다.

도 14b 는 본 개시의 일부 예들에 따른, 도 1 의 디바이스에 의해 수행될 수도 있는 자기 음성 활동 검출의 방법의 또 다른 특정 구현의 다이어그램이다.

도 15 은 본 개시의 일부 예들에 따른, 자기 음성 활동 검출을 수행하도록 동작가능한 디바이스의 특정한 예시적인 양태의 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0015] **VI. 상세한 설명**

[0016] 하나 이상의 컴포넌트들 또는 동작들의 활성화가 비-사용자 스피치로부터 기인하는 "오경보들"을 감소시키는 자기 음성 활동 검출 (self-voice activity detection, "SVAD") 은 오경보들이 검출될 때 그러한 컴포넌트들 또는 동작들의 활성화를 방지함으로써 디바이스의 전력 소비 효율을 개선할 수 있다. 그러나, SVAD 정확도를 개선하기 위한 종래의 오디오 신호 프로세싱 기법들은 또한 개선된 정확도 기술들을 수행하면서 디바이스의 전력 소비 및 프로세싱 리소스들을 증가시킨다. 디바이스가 저전력 또는 슬립 모드에 있는 동안에도, SVAD 프로세싱이 통상적으로 지속적으로 동작하기 때문에, 종래의 SVAD 기술들을 사용하여 오경보들을 감소시키는 것에 의해 전력 소비의 감소는 SVAD 프로세싱 자체와 연관된 증가된 전력 소비에 의해 부분적으로 또는 완전히 오프셋될 수 있다.

[0017] 동적 분류기를 사용하는 자기-음성 활동 검출의 시스템들 및 방법들이 개시된다. 예를 들어, 헤드셋 구현에서, 오디오 신호들은 사용자의 음성을 캡처하도록 위치설정되는 제 1 마이크로폰으로부터 및 노이즈 감소 및 에코 소거를 수행하는 것과 같은 외부 사운드들을 캡처하도록 위치설정되는 제 2 마이크로폰으로부터 수신될 수 있다. 오디오 신호들은 IPD들 (interaural phase differences) 및 IID들 (interaural intensity differences) 을 포함하는 주파수 도메인 특징 세트들을 추출하도록 프로세싱될 수 있다.

[0018] 동적 분류기는 추출된 주파수 도메인 특징 세트들을 프로세싱하고 특징 세트들의 분류를 표시하는 출력을 생성한다. 동적 분류기는 사용자 음성 활동에 대응하는 특징 세트들과 다른 오디오 활동에 대응하는 특징 세트들을 구별하기 위해 특징 데이터의 적응적 클러스터링 및 특징 데이터 공간의 2개의 가장 구별되는 카테고리들 사이의 결정 경계의 조정을 수행할 수 있다. 예시적인 예에서, 동적 분류기는 자기-조직화 맵들을 사용하여 구현된다.

[0019] 동적 분류기는 추출된 특징 세트들을 사용하는 구별이 다양한 조건들, 예를 들어 매우 비정적 상황들에서의 환경 조건들; 미스매칭된 마이크로폰들; 사용자 헤드셋 피팅에서의 변화들; 상이한 사용자 헤드-관련 전달 함수들 ("HRTF들") ; 비-사용자 신호들의 도착 방향 ("DOA") 추적; 주파수 스펙트럼에 걸친 마이크로폰들의 노이즈 플로어, 바이어스, 및 민감도들; 또는 그 조합에 능동적으로 응답하고 적응하는 것을 가능하게 한다. 일부 구현들에서, 동적 분류기는 이러한 변동들에 응답하고 사용된 임계화 파라미터들의 수 및 고객들에 의한 헤드셋 튜닝의 양을 감소시키거나 최소화할 수 있는 적응성 특징 맵핑을 가능하게 한다. 일부 구현예들에서, 동적 분류기는, 필적하는 정확도를 제공하는 종래의 SVAD 시스템들에 비해, 변하는 조건들 하에서 높은 정확도로 그리고 비교적 낮은 전력 소비로 사용자 음성 활동과 다른 오디오 활동 사이의 효과적인 구별을 가능하게 한다.

[0020] 본 개시의 특정 양태들은 도면들을 참조하여 이하에 설명된다. 설명에서, 공통 피처들 (features) 은 공통 참조 번호들로 지정된다. 본 명세서에서 사용된 바와 같이, 다양한 용어가 단지 특정 구현들을 설명할 목적으로 사용되고 구현들을 한정하는 것으로 의도되지 않는다. 예를 들어, 단수 형태들 "a", "an", 및 "the" 는, 문맥이 분명히 달리 표시하지 않는 한, 복수 형태들을 물론 포함하도록 의도된다. 또한, 본 명세서에 설명된 일부 특징은 일부 구현에서 단수이고 다른 구현에서 복수이다. 예시를 위해, 도 1 은 하나 이상의 프로세서들 (도 1 의 "프로세서 (들)" (190)) 을 포함하는 디바이스 (102) 를 도시하며, 이는 일부 구현들에서,

디바이스 (102) 가 단일 프로세서 (190) 를 포함하고, 다른 구현들에서, 디바이스 (102) 가 다수의 프로세서들 (190) 을 포함한다는 것을 표시한다. 본 명세서에서의 참조의 용이함을 위해, 이러한 특징들은 일반적으로 "하나 이상의" 특징들로서 도입되고, 후속적으로, 특징들 중 다수의 특징들과 관련된 양태들이 설명되지 않는 한 단수로 지칭된다.

[0021] 용어들 "포함한다 (comprise)", "포함한다 (comprises)", 및 "포함하는 (comprising)" 은 "포함한다 (include)", "포함한다 (includes)", 또는 "포함하는 (including)" 과 상호교환가능하게 사용될 수도 있는 것으로 추가로 이해될 수도 있다. 추가적으로, 용어 "여기서 (wherein)" 는 "여기에서 (where)" 와 상호교환가능하게 사용될 수도 있는 것으로 이해될 것이다. 본원에서 사용될 때, "예시적인" 은 예, 구현예, 및/또는 양태를 표시할 수도 있으며, 한정하거나 또는 선호사항 또는 바람직한 구현예를 표시하는 것으로 해석되어서는 안된다. 본 명세서에 사용된 바와 같이, 구조, 컴포넌트, 동작 등과 같은 엘리먼트를 수정하는데 사용되는 서수 용어 (예를 들어, "제 1", "제 2", "제 3" 등) 는 그것만으로 그 엘리먼트의 다른 엘리먼트에 대한 어떤 우선순위 또는 순서도 표시하지 않고, 오히려 그 엘리먼트를 (서수 용어의 사용이 없다면) 동일 명칭을 갖는 다른 엘리먼트와 구별할 뿐이다. 본원에서 사용될 때, 용어 "세트" 는 하나 이상의 특정의 엘리먼트를 지칭하며, 용어 "복수" 는 다수의 (예를 들어, 2개 이상의) 특정의 엘리먼트를 지칭한다.

[0022] 본원에 사용된 바와 같이, "커플링된" 은 "통신적으로 커플링된", "전기적으로 커플링된", 또는 "물리적으로 커플링된" 을 포함할 수도 있고, 또한 (또는 대안적으로) 이들의 임의의 조합을 포함할 수도 있다. 2 개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 다른 디바이스, 컴포넌트, 와이어, 버스, 네트워크 (예를 들어, 유선 네트워크, 무선 네트워크, 또는 그 조합) 등을 통해 직접적으로 또는 간접적으로 커플링될 (예를 들어, 통신적으로 커플링될, 전기적으로 커플링될, 또는 물리적으로 커플링될) 수도 있다. 전기적으로 커플링된 2 개의 디바이스들 (또는 컴포넌트들) 은 동일한 디바이스 또는 상이한 디바이스에 포함될 수도 있고, 예시적인, 비-제한적인 예들로서 전자기기들, 하나 이상의 커넥터들 또는 유도 결합을 통해 접속될 수도 있다. 일부 구현들에서, 전기 통신에서와 같이, 통신가능하게 커플링되는 2 개의 디바이스들 (또는 컴포넌트들) 은 하나 이상의 와이어들, 버스들, 네트워크들 등을 통해서, 직접 또는 간접적으로 신호들 (디지털 신호들 또는 아날로그 신호들) 을 전송 및 수신할 수도 있다. 본 명세서에서 사용된 바와 같이, "직접 커플링된" 은 컴포넌트들을 개재함 없이 커플링 (예를 들어, 통신가능하게 커플링, 전기적으로 커플링, 또는 물리적으로 커플링) 되는 2 개의 디바이스들을 포함할 수도 있다.

[0023] 본 개시물에서, 용어들 예를 들어 "결정하는 것", "계산하는 것", "추정하는 것", "시프트하는 것", "조정하는 것", 등은 하나 이상의 동작들이 수행되는 방법을 설명하기 위해 사용될 수도 있다. 이러한 용어들은 한정하는 것으로서 해석되지 않아야 하고 다른 기법들이 유사한 동작들을 수행하는데 활용될 수도 있음에 유의해야 한다. 추가적으로, 본 명세서에서 언급된 바와 같이, "생성하는 것", "계산하는 것", "추정하는 것", "사용하는 것", "선택하는 것", "액세스하는 것", 및 "결정하는 것" 은 상호교환가능하게 사용될 수도 있다. 예를 들어, 파라미터 (또는 신호) 를 "생성하는 것", "계산하는 것", “추정하는 것”, 또는 "결정하는 것" 은 파라미터 (또는 신호) 를 능동적으로 생성하는 것, 추정하는 것, 계산하는 것, 또는 결정하는 것을 지칭할 수도 있거나, 또는 예를 들어, 다른 컴포넌트 또는 디바이스에 의해 이미 생성된 파라미터 (또는 신호) 를 사용하는 것, 선택하는 것, 또는 액세스하는 것을 지칭할 수도 있다.

[0024] 도 1 을 참조하면, 동적 분류기를 사용하여 자기 음성 활동 검출을 수행하도록 구성된 시스템의 특정 예시적인 양태가 개시되고 일반적으로 100 으로 지정된다. 시스템 (100) 은 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 및 제 2 디바이스 (160) 에 커플링된 디바이스 (102) 를 포함한다. 디바이스 (102) 는 동적 분류기 (140) 를 사용하여 마이크로폰들 (110, 120) 에 의해 캡처된 사운드들의 자기-음성 활동 검출을 수행하도록 구성된다. 예시를 위해, 디바이스 (102) 가 헤드셋에 대응하는 구현에서, 제 1 마이크로폰 (110)(예를 들어, "1차" 마이크로폰) 은 디바이스 (102) 의 착용자의 입에 근접하게 위치설정된 마이크로폰과 같은 디바이스 (102) 의 사용자의 발화들을 1차적으로 캡처하도록 구성될 수도 있고, 제 2 마이크로폰 (120)(예를 들어, "2차" 마이크로폰) 은 착용자의 귀에 근접하게 위치설정된 것과 같은 주변 사운드를 1차적으로 캡처하도록 구성될 수도 있다. 디바이스 (102) 가 다수의 사람들 근처에 있을 수 있는 독립형 음성 어시스턴트 (예를 들어, 도 11 을 참조하여 추가로 설명되는 바와 같이 마이크로폰들을 갖는 라우드스피커를 포함함) 에 대응할 때와 같은 다른 구현들에서, 디바이스 (102) 는, 사람이 헤드셋 구현에 비해 1차 마이크로폰으로부터 비교적 멀리 있을 수 있더라도, 자기 음성 활동으로서 1차 마이크로폰에 가장 가까운 사람으로부터의 스피치를 검출하도록 구성될 수 있다. 본 명세서에서 사용되는 바와 같이, 용어 "자기-음성 활동 검출"은 디바이스의 사용자로부터 기원되지 않는 사운드들 (예를 들어, "다른 오디오 활동") 과 비교하여 디바이스 (102) 의 사용자의 스피치 (예를 들

어, 음성 또는 발화)(예를 들어, "사용자 음성 활동") 를 구별하는 것을 표시하기 위해 "사용자 음성 활동 검출"과 상호교환적으로 사용된다.

[0025] 디바이스 (102) 는 제 1 입력 인터페이스 (114), 제 2 입력 인터페이스 (124), 하나 이상의 프로세서들 (190), 및 모뎀 (170) 을 포함한다. 제 1 입력 인터페이스 (114) 는 프로세서 (190) 에 커플링되고, 제 1 마이크로폰 (110) 에 커플링되도록 구성된다. 제 1 입력 인터페이스 (114) 는 제 1 마이크로폰 (110) 으로부터 제 1 마이크로폰 출력 (112) 을 수신하고 제 1 마이크로폰 출력 (112) 을 제 1 오디오 데이터 (116) 로서 프로세서 (190) 에 제공하도록 구성된다.

[0026] 제 2 입력 인터페이스 (124) 는 프로세서 (190) 에 커플링되고, 제 2 마이크 (120) 에 커플링되도록 구성된다. 제 2 입력 인터페이스 (124) 는 제 2 마이크로폰 (120) 으로부터 제 2 마이크로폰 출력 (122) 을 수신하고 제 2 마이크로폰 출력 (122) 을 제 2 오디오 데이터 (126) 로서 프로세서 (190) 에 제공하도록 구성된다.

[0027] 프로세서 (190) 는 모뎀 (170) 에 커플링되고, 특징 추출기 (130) 및 동적 분류기 (140) 를 포함한다. 프로세서는 제 1 마이크로폰 (110) 의 제 1 출력 (112) 에 대응하는 제 1 오디오 데이터 (116) 및 제 2 마이크로폰 (120) 의 제 2 출력 (122) 에 대응하는 제 2 오디오 데이터 (126) 를 포함하는 오디오 데이터 (128) 를 수신하도록 구성된다. 프로세서 (190) 는 특징 추출기 (130) 에서 오디오 데이터 (128) 를 프로세싱하여 특징 데이터 (132) 를 생성하도록 구성된다.

[0028] 일부 구현들에서, 프로세서 (190) 는 특징 데이터 (132) 를 생성하기 전에 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 프로세싱하도록 구성된다. 일 예에서, 프로세서 (190) 는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 에 대해 에코-소거, 노이즈 억제, 또는 양자 모두를 수행하도록 구성된다. 일부 구현들에서, 프로세서 (190) 는 특징 데이터 (132) 를 생성하기 전에 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 변환 도메인으로 변환 (예를 들어, 푸리에 변환) 하도록 구성된다.

[0029] 프로세서 (190) 는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 에 기초하여 특징 데이터 (132) 를 생성하도록 구성된다. 일부 양태들에 따르면, 특징 데이터 (132) 는 제 1 오디오 데이터 (116) 와 제 2 오디오 데이터 (126) 사이의 적어도 하나의 귀간 위상 차이 (134) 및 제 1 오디오 데이터 (116) 와 제 2 오디오 데이터 (126) 사이의 적어도 하나의 귀간 강도 차이 (136) 를 포함한다. 특정 예에서, 특징 데이터 (132) 는 다수의 주파수들에 대한 귀간 위상 차이들 (IPD들)(134) 및 다수의 주파수들에 대한 귀간 강도 차이들 (IID들)(136) 을 포함한다.

[0030] 프로세서 (190) 는 특징 데이터 (132) 의 분류 출력 (142) 을 생성하기 위해 동적 분류기 (140) 에서 특징 데이터 (132) 를 프로세싱하도록 구성된다. 일부 구현들에서, 동적 분류기 (140) 는 오디오 데이터 (128) 에 표현된 사운드가 제 2 마이크로폰 (120) 보다 제 1 마이크로폰 (110) 에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터 (132) 의 세트들 (예를 들어, 샘플들) 을 적응적으로 클러스터링하도록 구성된다. 예를 들어, 동적 분류기 (140) 는 특징 데이터 (132) 의 샘플들의 시퀀스를 수신하고 IID 및 IPD 주파수 값들을 포함하는 특징 공간에서 샘플들을 적응적으로 클러스터링하도록 구성될 수도 있다.

[0031] 동적 분류기 (140) 는 또한 사용자 음성 활동 (예를 들어, 사용자 (180) 의 발화 (182)) 에 대응하는 특징 데이터의 세트들과 다른 오디오 활동에 대응하는 특징 데이터의 세트들 사이를 구별하기 위해 특징 공간의 2개의 가장 구별되는 카테고리들 사이의 결정 경계를 조정하도록 구성될 수 있다. 예시를 위해, 동적 분류기 (140) 는 인커밍 특징 데이터를 2개의 클래스 (예를 들어, 클래스 0 또는 클래스 1) 중 하나로 분류하도록 구성될 수 있으며, 여기서 2개의 클래스 중 하나는 사용자 음성 활동에 대응하고, 2개의 클래스 중 다른 하나는 다른 오디오 활동에 대응한다. 분류 출력 (142) 은 2개의 값들 중 하나, 즉 특징 데이터 (132) 가 2개의 클래스들 중 하나에 대응함을 표시하는 제 1 값 (예를 들어, "0") ; 또는 특징 데이터 (132) 가 2개의 클래스들 중 다른 하나에 대응함을 나타내는 제 2 값 (예를 들어, "1") 을 갖는 단일 비트 또는 플래그를 포함할 수 있다.

[0032] 일부 구현들에서, 동적 분류기 (140) 는 클러스터링 및 벡터 양자화를 수행한다. 예를 들어, 클러스터링은

$$min \sum_{i=1}^{n} \sum_{x_j \in C_i} p_i \left\| \boldsymbol{x}_j - \boldsymbol{\mu}_i \right\|^2$$

과 같이 정의된 정사각형들의 클러스터 내 합 (within-cluster sum) 을 감소시키는 것 (예를 들어, 최소화하는 것) 을 포함하고, 여기서, C_i 는 클러스터 i 를 나타내며, p_i 는 클러스터 i 를 나타내고, $\boldsymbol{x}_j$ 는 특징 공간에서 노드 j 를 나타내고, $\boldsymbol{\mu}_i$ 는 클러스터 i 의 도심을 나타낸다. 클러스터 가중치 p_i 는 이전의 클러스터 분포와 같이 확률적일 수 있거나; 각각의 클러스터의 가능성에 할당된 신뢰 척도와

같이 확률적일 수 있거나; 또는 상이한 클러스터들를 향한 불균일한 바이어스의 일부 형태를 강제할 임의의 다른 인자에 의해 결정될 수 있다. 벡터 양자화는 입력 벡터를 $min \sum_{i=1}^{n} \sum_{x_j \in C_i} p_i \|x_j - w_i\|^2$ 로 정의된 양자화 가중치 벡터로 양자화함으로써 에러를 감소 (예를 들어, 최소화) 하는 것을 포함하며, 여기서 w_i 는 양자화 가중치 벡터 i 를 나타낸다.

[0033] 일부 구현들에서, 동적 분류기 (140) 는 양자화 유닛들이 특징 데이터 (132) 의 새로운 샘플들을 흡수하기 위해 경쟁하는 경쟁 학습을 수행하도록 구성된다. 그후, 승리 유닛은 새로운 샘플의 방향으로 조정된다. 예를 들어, 각각의 유닛의 가중치 벡터는 분리를 위해 초기화되거나 랜덤하게 초기화될 수 있다. 수신된 특징 데이터의 각각의 새로운 샘플에 대해, 비제한적인 예로서, 유클리드 거리 또는 내적 유사성 (inner product similarity) 에 기초하는 것과 같이, 어느 가중치 벡터가 새로운 샘플에 가장 가까운지에 관한 결정이 이루어진다. 그후, 새로운 샘플 ("승자" 또는 최상의 매칭 유닛) 에 가장 가까운 가중치 벡터가 새로운 샘플의 방향으로 이동될 수 있다. 예를 들어, Hebbian 학습에서 승자들은 예를 들어 2개의 노드들에 대한 입력들의 곱에 비례하여 2개의 노드들 사이의 가중치들을 조정함으로써 입력과의 그 상관들을 강화한다.

[0034] 일부 구현들에서, 동적 분류기 (140) 는 포스트시냅틱 시트에서의 로컬 클러스터들에 연결된 포스트시냅틱 시트에서의 로컬 클러스터들을 포함하고, 그리고 이웃 뉴런들 사이의 상호연결들은 상관 자극들 사이의 연결들을 강화하기 위해 Hebbian 학습을 통해 강화된다. 동적 분류기 (140) 는 포스트시냅틱 시트 또는 맵에서의 모든 뉴런에 입력이 연결된 Kohonen 자기 조직화 맵을 포함할 수 있다. 학습은 상이한 흡수 필드들이 입력 공간의 상이한 영역들 (예를 들어, 특징 데이터 공간) 에 대해 응답한다는 점에서 맵이 로컬화되게 한다.

[0035] 특정 구현에서, 동적 분류기 (140) 는 자기-조직화 맵 (148) 을 포함한다. 자기-조직화 맵 (140) 은 가중치 벡터들을 초기화하고, 그후 각각의 입력 t (예를 들어, 특징 데이터 (132) 의 각각의 수신된 세트) 에 대해,

$$v(t) = \arg\min_i \|x(t) - w_i(t)\|$$

에 따라 승리 유닛 (또는 셀 또는 뉴런) 을 결정함으로써, 승자 $v(t)$ 를 입력 $x(t)$ 에 대한 가장 작은 거리 (예를 들어, Euclidean 거리) 를 갖는 유닛으로서 발견할 수 있도록 동작될 수 있다. 승리 유닛 및 그 이웃들의 가중치는 예를 들어

$$\Delta w_i(t) = \alpha(t) l(v, i, t)[x(t) - w_v(t)]$$

에 따라서 업데이트되고, 여기서 $\Delta w_i(t)$ 는 유닛 i 에 대한 변경을 나타내고, $\alpha(t)$ 는 학습 파라미터를 나타내고, $l(v, i, t)$ 는 예를 들어 Gaussian 방사형 기저 함수로서 승리 유닛 주위의 이웃 함수를 나타낸다. 일부 구현들에서, 내부 곱들 또는 다른 메트릭이 Euclidean 거리 대신에 유사성 측정치로서 사용될 수 있다.

[0036] 일부 구현들에서, 동적 분류기 (140) 는 도 4 를 참조하여 추가로 설명되는 바와 같이, 스피치 샘플들의 시퀀스들을 수용하기 위한 Kohonen 자기-조직화 맵의 변형을 포함한다. 일 실시예에서, 동적 분류기 (140) 는 예를 들어 시간 상수 모델링 감쇠 (time constant modeling decay, D) 를 갖는 활성화 함수가 각각의 유닛에 대해 정의되고 $U_i(t, D) = D.U_i(t-1, D) - \frac{1}{2}\|x(t) - w_i(t)\|^2$ 로서 업데이트되고 승리 유닛이 가장 큰 활동을 갖는 유닛인 시간 Kohonen 맵에 따라 시간 시퀀스 프로세싱을 구현할 수 있다. 또 다른 예로, 동적 분류기 (140) 는 예를 들어 제곱 표준 (squared norm) : $y_i(t, \gamma) = (1-\gamma)y_i(t-1, \gamma) + \gamma(x(t) - w_i(t))$ 대신에 차분 벡터 (difference vector : y) 를 사용하는 반복적인 자기조직화 맵 (recurrent self-organizing map) 에 따라 반복 네트워크 (recurrent network) 를 구현할 수 있고, 여기서, γ 는 0 내지 1 사이의 값을 갖는 망각 팩터 (forgetting factor) 를 나타내며, 승리 유닛은 가장 작은 차분 벡터 $v(t) = \arg\min_i \|y_i(t, \gamma)\|$ 의 유닛으로서 결정되고, 가중치는 $\Delta w_i(t) = \alpha(t) l(v, i, t)[x(t) - y_v(t, \gamma)]$ 로서 업데이트된다.

[0037] 일부 구현들에서, 프로세서 (190) 는 특징 데이터 (132) 에 기초하여 동적 분류기 (140) 의 클러스터링 동작 (144) 을 업데이트하고 동적 분류기 (140) 의 분류 결정 기준 (146) 을 업데이트하도록 구성된다. 예를 들어, 앞서 설명된 바와 같이, 프로세서 (190) 는 오디오 데이터 (128) 의 인커밍 샘플들에 기초하여 사용자 음성 활동과 다른 오디오 활동 사이의 클러스터링 및 결정 경계를 적응시키도록 구성되어, 동적 분류기 (140) 가 사용자 (180) 의 변화하는 조건들, 환경, 다른 조건들 (예를 들어, 마이크로폰 배치 또는 조정), 또는 그 임의의

조합에 기초하여 동작을 조정할 수 있게 한다.

[0038] 동적 분류기 (140) 가 자기-조직화 맵 (148) 을 포함하는 것으로 예시되지만, 다른 구현들에서, 동적 분류기 (140) 는 자기-조직화 맵 (148) 대신에 또는 그에 부가하여 분류 출력 (142) 을 생성하기 위해 하나 이상의 다른 기술들을 통합할 수 있다. 비제한적인 예로서, 동적 분류기 (140) 는 비지도 (unsupervise) 구성을 갖는 제한된 볼츠만 머신, 비지도 오토인코더, Hopfield 네트워크의 온라인 변형, 온라인 클러스터링, 또는 그 조합을 포함할 수 있다. 다른 비제한적인 예로서, 동적 분류기 (140) 는 주 컴포넌트 분석 (principal component analysis) 을 수행하도록 구성될 수 있다 (예를 들어, 설정된 직교 방향 벡터들을 특징 공간에서 특징 벡터 샘플들에 순차적으로 피팅하고, 여기서 각각의 방향 벡터는 특징 공간에서 방향 벡터 상에 투영된 특징 벡터 샘플들의 분산을 최대화하는 것으로서 선택된다). 다른 비제한적인 예로서, 동적 분류기 (140) 는 독립적인 컴포넌트 분석 (예를 들어, 서브컴포넌트들이 서로 통계적으로 독립적인 비-Gaussian 신호들이라는 가정하에, 특징 공간에서 특징 벡터 샘플들의 부가 서브컴포넌트들의 세트를 결정함) 을 수행하도록 구성될 수도 있다.

[0039] 프로세서 (190) 는, 분류 출력 (142) 에 적어도 부분적으로 기초하여, 오디오 데이터 (128) 가 사용자 음성 활동에 대응하는지 여부를 결정하고, 사용자 음성 활동이 검출되는지 여부를 표시하는 사용자 음성 활동 표시기 (150) 를 생성하도록 구성된다. 예를 들어, 분류 출력 (142) 은 특징 데이터 (132) 가 2개의 클래스 (예를 들어, 클래스 "0" 또는 클래스 "1") 중 하나로서 분류되는지 여부를 표시할 수 있지만, 분류 출력 (142) 은 어느 클래스가 사용자 음성 활동에 대응하고 어느 클래스가 다른 오디오 활동에 대응하는지를 표시하지 않을 수 있다. 예를 들어, 동적 분류기 (140) 가 어떻게 초기화되는지 및 동적 분류기 (140) 를 업데이트하는데 사용된 특징 데이터에 기초하여, 일부 경우들에서 값 "0" 을 갖는 분류 출력 (142) 은 사용자 음성 활동을 표시하는 반면, 다른 경우들에서 값 "0" 을 갖는 분류 출력은 다른 오디오 활동을 표시한다. 프로세서 (190) 는 도 2 를 참조하여 추가로 설명되는 바와 같이, 특징 데이터 (132) 의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초하여, 2개의 클래스 중 어느 것이 사용자 음성 활동을 표시하는지 및 2개의 클래스 중 어느 것이 다른 오디오 활동을 표시하는지를 결정할 수 있다.

[0040] 예시를 위해, 사용자 (180) 의 입으로부터 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 으로의 발화 (182) 의 사운드 전파는 (발화 (182) 가 제 2 마이크로폰 (120) 이전에 제 1 마이크로폰 (110) 에 도달함으로 인해) 위상 차이 및 특징 데이터 (132) 에서 검출될 수 있고 다른 오디오 소스들로부터의 사운드의 위상 및 신호 강도 차이들과 구별가능할 수 있는 신호 강도 차이를 초래한다. 위상 및 신호 강도 차이들은 특징 데이터 (132) 에서 IPD들 (134) 및 IID들 (136) 로부터 결정될 수 있고, 분류 출력 (142) 을 사용자 음성 활동 또는 다른 오디오 활동에 맵핑하는데 사용될 수 있다. 프로세서 (190) 는 오디오 데이터 (128) 가 사용자 음성 활동에 대응하는지 여부를 표시하는 사용자 음성 활동 표시기 (150) 를 생성할 수 있다.

[0041] 일부 구현들에서, 프로세서 (190) 는 오디오 데이터 (128) 가 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작 (152) 을 개시하도록 구성된다. 예시적인 예에서, 음성 커맨드 프로세싱 동작 (152) 은 키워드 또는 키 구문 검출, 음성 프린트 인증, 자연어 프로세싱, 하나 이상의 다른 동작, 또는 그 임의의 조합과 같은 음성 활성화 동작을 포함한다. 또 다른 예로서, 프로세서 (190) 는 키워드 검출의 제 1 스테이지를 수행하기 위해 오디오 데이터 (128) 를 프로세싱할 수 있고, (예를 들어, 더 강력한 음성 활동 인식 및 스피치 인식 동작들을 포함하는 검출의 제 2 스테이지에서) 음성 커맨드 프로세싱 동작 (152) 을 통해 오디오 데이터 (128) 의 추가 프로세싱을 개시하기 전에, 검출된 키워드가 인근 사람에 의해서라기보다는 디바이스 (102) 의 사용자 (180) 에 의해 스피치되었다는 것을 확인하기 위해 사용자 음성 활동 표시기 (150) 를 사용할 수 있다.

[0042] 모뎀 (170) 은 프로세서 (190) 에 커플링되고, 예를 들어, 무선 송신을 통해 제 2 디바이스 (160) 와의 통신을 가능하게 하도록 구성된다. 일부 예들에서, 모뎀 (170) 은 오디오 데이터 (128) 가 동적 분류기 (140) 에 기초한 사용자 음성 활동에 대응한다는 결정에 응답하여 오디오 데이터 (128) 를 제 2 디바이스 (160) 로 송신하도록 구성된다. 예를 들어, 디바이스 (102) 가 제 2 디바이스 (160) 에 무선으로 커플링되는 헤드셋 디바이스 (예를 들어, 모바일 폰 또는 컴퓨터에 대한 블루투스 접속) 에 대응하는 구현에서, 디바이스 (102) 는 제 2 디바이스 (160) 의 음성 활성화 시스템 (162) 에서 음성 커맨드 프로세싱 동작 (152) 을 수행하기 위해 오디오 데이터 (128) 를 제 2 디바이스 (160) 로 전송할 수도 있다. 이 예에서, 디바이스 (102) 는 제 2 디바이스 (160) 의 더 큰 프로세싱 리소스들 및 전력 리소스들을 사용하여 수행될 더 계산적으로 비싼 프로세싱 (예를 들어, 음성 커맨드 프로세싱 동작 (152)) 을 오프로드한다.

[0043] 일부 구현들에서, 디바이스 (102) 는 하나 또는 다양한 타입들의 디바이스들에 대응하거나 그에 포함된다. 예시적인 예에서, 프로세서 (190) 는 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 을 포함하는 헤드셋 디바이스에 통합된다. 헤드셋 디바이스는, 사용자 (180) 에 의해 착용될 때, 도 7 을 참조하여 추가로 설명되는 바와 같이, 제 2 마이크로폰 (120) 에서와 비교되는 바와 같이 보다 큰 강도 및 더 늦은 지연으로 제 1 마이크로폰 (110) 에서 사용자 (180) 의 발화들 (182) 을 캡처하기 위해 사용자의 입에 제 2 마이크로폰 (120) 보다 더 가깝게 제 1 마이크로폰 (110) 을 위치설정하도록 구성된다. 다른 예들에서, 프로세서 (190) 는 도 6 을 참조하여 설명된 바와 같이, 모바일 폰 또는 태블릿 컴퓨터 디바이스, 도 8 을 참조하여 설명된 바와 같이, 웨어러블 전자 디바이스, 도 9 를 참조하여 설명된 바와 같이, 음성 제어 스피커 시스템, 도 10 을 참조하여 설명된 바와 같이, 카메라 디바이스, 또는 도 11 을 참조하여 설명된 바와 같이, 가상 현실 헤드셋, 혼합 현실 헤드셋, 또는 증강 현실 헤드셋 중 적어도 하나에 통합된다. 또 다른 예시적인 예에서, 프로세서 (190) 는, 도 12 및 도 13 을 참조하여 추가로 설명되는 바와 같이, 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 을 또한 포함하는 비히클에 통합된다.

[0044] 동작 동안, 제 1 마이크로폰 (110) 은 사용자 (180) 의 발화들 (182) 을 캡처하도록 구성되고, 제 2 마이크로폰 (120) 은 주변 사운드 (186) 를 캡처하도록 구성된다. 일 예에서, 디바이스 (102) 의 사용자 (180) 로부터의 발화 (182) 는 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 에 의해 캡처된다. 제 1 마이크로폰 (110) 이 사용자 (180) 의 입에 더 가깝기 때문에, 사용자 (180) 의 스피치는 제 2 마이크로폰 (120) 에 비해 더 높은 신호 강도 및 더 적은 지연으로 제 1 마이크로폰 (110) 에 의해 캡처된다. 또 다른 예에서, 하나 이상의 사운드 소스들 (184) 로부터의 주변 사운드 (186)(예를 들어, 2 명의 근처의 사람들 사이의 대화) 는 제 1 마이크로폰 (110) 에 의해 및 제 2 마이크로폰 (120) 에 의해 캡처될 수도 있다. 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 에 대한 사운드 소스들 (184) 의 포지션 및 거리에 기초하여, 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 에서 주변 사운드 (186) 를 캡처하는 것 사이의 신호 강도 차이 및 상대 지연은 사용자 (180) 로부터의 발화 (182) 에 대한 것과 다를 것이다.

[0045] 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 는 에코 소거, 노이즈 억제, 주파수 도메인 변환 등을 수행함으로써 프로세서 (190) 에서 프로세싱된다. 결과적인 오디오 데이터는 IPD들 (134) 및 IID들 (136) 을 포함하는 특징 데이터 (132) 를 생성하기 위해 특징 추출기 (130) 에서 프로세싱된다. 특징 데이터 (132) 는 프로세서 (190) 에 의해 사용자 음성 활동 또는 다른 사운드 활동으로서 해석되는, 분류 출력 (142) 을 생성하기 위해 동적 분류기 (140) 에 입력된다. 프로세서 (190) 는 오디오 데이터 (128) 가 사용자 음성 활동에 대응한다는 것을 표시하는 "0" 값, 또는 오디오 데이터 (128) 가 다른 오디오 활동에 대응한다는 것을 나타내는 "1" 값 (또는 그 반대) 과 같은 사용자 음성 활동 표시기 (150) 를 생성한다.

[0046] 사용자 음성 활동 표시기 (150) 는 디바이스 (102) 에서 음성 커맨드 프로세싱 동작 (152) 을 개시할지 여부를 결정하는 데 사용될 수 있다. 대안적으로 또는 추가적으로, 사용자 음성 활동 표시기 (150) 는 음성 활성화 시스템 (162) 에서의 추가 프로세싱을 위해 제 2 디바이스 (160) 에 대한 출력 신호 (135)(예를 들어, 오디오 데이터 (128)) 의 생성을 개시할지 여부를 결정하기 위해 사용될 수 있다.

[0047] 또한, 분류 출력 (142) 을 생성하는 것과 함께, 동적 분류기 (140) 는 특징 데이터 (132) 에 기초하여, 예를 들어 클러스터링 동작 (144), 분류 기준 (146), 또는 그 조합을 업데이트하는, 승리 유닛 및 그 이웃들의 가중치들을 특징 데이터 (132) 와 더 유사하도록 조정함으로써, 업데이트된다. 이러한 방식으로, 동적 분류기 (140) 는 사용자 스피치의 변화들, 환경의 변화들, 디바이스 (102) 또는 마이크로폰들 (110, 120) 의 특성들의 변화들, 또는 그 조합에 자동으로 적응한다.

[0048] 따라서, 시스템 (100) 은 종래의 자기-음성 활동 검출 기술들에 비해 상대적으로 낮은 복잡도, 낮은 전력 소비, 및 높은 정확도로 사용자 음성 활동과 다른 오디오 활동 사이에서 구별하기 위해 동적 분류기 (140) 를 사용함으로써 자기-음성 활동 검출의 성능을 개선한다. 사용자 및 환경 변화에 자동으로 적응하는 것은 사용자에 의해 수행되는 캘리브레이션을 감소시키거나 제거함으로써 개선된 이점을 제공하고 사용자의 경험을 향상시킨다.

[0049] 일부 구현들에서 프로세서 (190) 는 특징 추출기 (130) 에 의해 생성되는 특징 데이터 (132)(예를 들어, 주파수 도메인 데이터) 의 형태로 오디오 데이터 (128) 를 동적 분류기 (140) 에 제공하지만, 다른 구현들에서 특징 추출기 (130) 는 생략된다. 일 예에서, 프로세서 (190) 는 오디오 데이터 (128) 를 오디오 샘플들의 시계열로서 동적 분류기 (140) 에 제공하고, 동적 분류기 (140) 는 분류 출력 (142) 을 생성하기 위해 오디오 데이터 (128) 를 프로세싱한다. 예시적인 구현에서, 동적 분류기 (140) 는 오디오 데이터 (128) 로부터 주파수 도

메인 데이터를 결정하고 (예를 들어, 특징 데이터 (132) 를 생성하고) 추출된 주파수 도메인 데이터를 사용하여 분류 출력 (142) 을 생성하도록 구성된다.

[0050] 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 이 디바이스 (102) 에 커플링되는 것으로서 예시되지만, 다른 구현들에서, 제 1 마이크로폰 (110) 또는 제 2 마이크로폰 (120) 중 하나 또는 양자는 디바이스 (102) 에 통합될 수도 있다. 2개의 마이크로폰들 (110, 120) 이 예시되지만, 다른 구현들에서 사용자 스피치를 캡처하도록 구성된 하나 이상의 추가적인 마이크로폰들, 환경 사운드들을 캡처하도록 구성된 하나 이상의 마이크로폰들, 또는 양자 모두가 포함될 수도 있다. 시스템 (100) 이 제 2 디바이스 (160) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 제 2 디바이스 (160) 는 생략될 수도 있고, 디바이스 (102) 는 제 2 디바이스 (160) 에서 수행되는 바와 같은 설명된 동작들을 수행할 수도 있다.

[0051] 도 2 는 도 1 의 디바이스 (102)(예를 들어, 프로세서 (190)) 에 의해 수행될 수도 있는 자기 음성 활동 검출과 연관된 동작들 (200) 의 예시적인 양태의 다이어그램이다. 특징 추출 (204) 은 특징 데이터 (206) 를 생성하기 위해 입력 (202) 에 대해 수행된다. 일 예에서, 입력 (202) 은 오디오 데이터 (128) 에 대응하고, 특징 추출 (204) 은 특징 추출기 (130) 에 의해 수행되고, 특징 데이터 (206) 는 특징 데이터 (132) 에 대응한다.

[0052] 동적 분류기 (208) 는 분류 출력 (210) 을 생성하기 위해 특징 데이터 (206) 에 대해 동작한다. 일 예에서, 동적 분류기 (208) 는 동적 분류기 (140) 에 대응하고, 분류 출력 (210) 에서 음성 활성화 클래스들에 대한 "자기" 대 "다른" 라벨링에 대한 높은 동적 결정 경계들을 갖는 특징 데이터 (206) 에 기초하여 비지도 실시간 클러스터링을 수행하도록 구성된다. 예를 들어, 동적 분류기 (208) 는 특징 공간을 2개의 클래스로 분할할 수 있는데, 하나의 클래스는 사용자 음성 활동과 연관되고 다른 클래스는 다른 사운드 활동과 연관된다. 분류 출력 (210) 은 클래스가 특징 데이터 (206) 와 연관되는 이진 표시기를 포함할 수 있다. 일 예에서, 분류 출력 (210) 은 분류 출력 (142) 에 대응한다.

[0053] 자기/다른 연관 동작 (212) 은 분류 출력 (210) 및 검증 입력 (216) 에 기초하여 자기/다른 표시기 (218) 를 생성한다. 검증 입력 (216) 은 분류 출력 (210) 의 클래스들 각각을 사용자 음성 활동 (예를 들어, "자기") 또는 다른 사운드 활동 (예를 들어, "다른 것") 과 연관시키는 정보를 제공할 수 있다. 예를 들어, 검증 입력 (216) 은 위상 차이의 부호 (230)(예를 들어, 어느 마이크로폰이 입력 (202) 에 의해 표현된 오디오의 소스에 더 가까운지를 표시하는, 하나 이상의 특정 주파수 범위들에 걸친 IPD들 (134) 중 하나 이상의 값) 를 비교하는 것, 강도 차이의 크기 (232) 를 비교하는 것 (예를 들어, 분리된 마이크로폰들에 대한 오디오의 소스의 상대적인 거리들을 표시하는, 하나 이상의 특정 주파수 범위들에 걸친 IID들 (136) 중 하나 이상의 값), 또는 그 조합과 같은 적어도 하나의 사전 검증 기준 (214) 에 기초하여 생성될 수도 있다. 예를 들어, 자기/다른 연관성은 "0"의 분류 출력 (210) 값이 하나 이상의 관련 주파수 범위들에서 음의 부호 (230) 를 나타내거나, 하나 이상의 관련 주파수 범위들에서 임계량보다 작은 크기 (232) 를 나타내거나, 또는 둘 모두를 나타내는 특징 데이터 (206) 에 대응한다는 것을 결정할 수 있고, 그 결과 "0"이 "다른 것"에 대응하고 "1"이 "자기"에 대응하도록 테이블을 채울 수 있다.

[0054] 자기/다른 연관 동작 (212) 은 자기/다른 표시기 (218)(예를 들어, 사용자 음성 활동을 표시하기 위해 제 1 값 (예를 들어, "0") 을 갖거나 다른 사운드 활동을 표시하기 위해 제 2 값 (예를 들어, "1"), 또는 그 반대를 갖는 이진 표시기) 의 생성을 초래한다. 웨이크업/바지-인 (barge-in) 제어 동작 (220) 은 음성 커맨드 프로세스 (224) 에 대한 신호 (222) 를 생성하기 위해 자기/다른 표시기 (218) 에 응답한다. 예를 들어, 신호 (222) 는 음성 명령 프로세스 (224) 가 입력 (202), 특징 데이터 (206) 또는 둘 모두에 대해 실행되는 것을 표시하여, 입력 (202) 이 사용자 음성 활동에 대응할 때 추가 음성 커맨드 프로세스를 수행하기 위해 (예를 들어, 키워드 검출, 음성 인증 또는 둘 모두를 수행하기 위해) 제 1 값 (예를 들어, "0") 을 가질 수 있거나, 또는 입력 (202) 이 다른 사운드 활동에 대응할 때 음성 커맨드 프로세스 (224) 가 음성 커맨드 프로세싱을 수행하지 않는다는 것을 표시하기 위해 제 2 값 (예를 들어, "1") 을 가질 수 있다.

[0055] 도 1 의 동적 분류기 (140) 및 도 2의 동적 분류기 (208) 를 참조하여 설명된 바와 같은 동적 분류는, 사용자가 스피치할 때만 응답하고 다른 간섭들 (예를 들어, 외부 음성) 이 도달할 때 항상 응답들을 억제하고, 자기 키워드 수용율 ("SKAR") 및 다른 키워드 거절율 ("OKRR") 을 최대화하는 것을 목적으로 하는, SVAD 정확도를 개선하는 것을 돕는다. 동적 분류를 사용함으로써, 종래의 SVAD 프로세싱과 연관된 다양한 도전들이 회피되거나 그렇지 않으면 감소된다. 예를 들어, 동적 분류의 구현을 통해 회피되거나 감소되는 종래의 SVAD 프로세싱 도전과제들은 (가혹한 조건들 하에서 잘못된 웨이크업 및 바지-인을 야기할 수 있는) 노이즈 및 에코 조건들, 마이크로폰 불일치 및 민감도, 음성 활성화 엔진 종속성, 상이한 사용자 헤드-관련 전달 함수들 (HRTF

들), 상이한 헤드셋 하드웨어 효과들, 폐색 및 격리 레벨들의 사용자 거동 구동 변동들, 다른 음성 활동의 사용자 특징 유사성, 음성 활성화에 대한 최종 부정적 효과들, 및 사용자 스피치의 시작의 응답 지연을 포함한다.

예시를 위해, 종래의 SVAD 는 내부/외부 마이크로폰 캘리브레이션 및 민감도, 간섭 스피치의 도달 방향, 헤드셋 피팅 및 격리의 변화, 및 동적 분류의 동작에 의해 수용될 수 있는 특징들의 비정적 통계에 크게 의존한다.

[0056] 동적 분류의 사용은 추출된 특징 데이터들 (206) 을 사용하는 구별이 다양한 조건들, 예를 들어 매우 비정적 상황들에서의 환경 조건들; 미스매칭된 마이크로폰들; 사용자 헤드셋 피팅에서의 변화들; 상이한 사용자 헤드-관련 전달 함수들; 비-사용자 신호들의 도착 방향 ("DOA") 추적; 및 주파수 스펙트럼에 걸친 마이크로폰들의 노이즈 플로어, 바이어스, 및 민감도들에 능동적으로 응답하고 적응하는 것을 가능하게 한다. 동적 분류는 이러한 변동들에 응답하고 사용된 임계화 파라미터들의 수 및 고객들에 의한 헤드셋 튜닝의 양을 감소시키거나 최소화할 수 있는 적응성 특징 맵핑을 가능하게 한다.

[0057] 도 3 은 본 개시의 일부 예들에 따른, 자기-음성 활동 검출을 수행하도록 동작가능한 시스템의 예시적인 양태의 블록도이며, 여기서 프로세서 (190) 는 올웨이즈-온 전력 도메인 (303) 및 온-디맨드 전력 도메인과 같은 제 2 전력 도메인 (305) 을 포함한다. 일부 구현들에서, 자기-음성 활동 검출기 (320) 의 제 1 스테이지 (340) 및 버퍼 (360) 는 올웨이즈-온 모드에서 동작하도록 구성되고, 자기-음성 활동 검출기 (320) 의 제 2 스테이지 (350) 는 온-디맨드 모드에서 동작하도록 구성된다.

[0058] 올웨이즈-온 도메인 (303) 은 버퍼 (360), 특징 추출기 (130) 및 동적 분류기 (140) 를 포함한다. 버퍼 (360) 는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 저장하여 자기-음성활동 검출기 (320) 의 컴포넌트들에 의한 프로세싱을 위해 액세스가능하도록 구성된다.

[0059] 제 2 전력 도메인 (305) 은 자기-음성 활동 검출기 (320) 의 제 2 스테이지 (250) 에 음성 커맨드 프로세싱 유닛 (370) 을 포함하고, 또한 활성화 회로 (330) 를 포함한다. 일부 구현예에서, 음성 커맨드 프로세싱 유닛 (370) 은 도 1 의 음성 커맨드 프로세싱 유닛 동작 (152) 또는 도 2 의 음성 커맨드 유닛 프로세스 (224) 를 수행하도록 구성된다.

[0060] 자기-음성 활동 검출기 (320) 의 제 1 스테이지 (240) 는 음성 커맨드 프로세싱 유닛 (370) 에서 음성 커맨드 프로세싱 동작 (152)(또는 음성 커맨드 프로세스 (224)) 을 개시하기 위해 웨이크업 신호 (322) 또는 인터럽트 (324) 중 적어도 하나를 생성하도록 구성된다. 일 예에서, 웨이크업 신호 (322) 는 음성 커맨드 프로세싱 유닛 (370) 을 활성화하기 위해 제 2 전력 도메인 (305) 을 저전력 모드 (332) 로부터 활성 모드 (334) 로 전환하도록 구성된다. 일부 구현들에서, 웨이크업 신호 (322), 인터럽트 (324), 또는 양자 모두는 도 2 의 신호 (222) 에 대응한다.

[0061] 예를 들어, 활성화 회로부 (330) 는 전력 관리 회로부, 클록 회로부, 헤드 스위치 또는 풋 스위치 회로부, 버퍼 제어 회로부, 또는 이들의 임의의 조합을 포함하거나 또는 그에 커플링될 수도 있다. 활성화 회로부 (330) 는, 예를 들어 제 2 스테이지 (350) 의, 제 2 전력 도메인 (305) 의, 또는 양자 모두의 전력 공급부의 전압을 선택적으로 인가 또는 상승시킴으로써, 제 2 스테이지 (350) 의 파워-온을 개시하도록 구성될 수도 있다. 다른 예로서, 활성화 회로부 (330) 는, 예를 들어, 전력 공급부를 제거하지 않고 회로 동작을 방지하거나 가능하게 하기 위해, 클록 신호를 제 2 스테이지 (350) 에 선택적으로 게이팅 또는 언게이트팅하도록 구성될 수 있다.

[0062] 자기-음성 활동 검출기 (320) 의 제 2 스테이지 (350) 에 의해 생성된 검출기 출력 (352) 은 애플리케이션 (354) 에 제공된다. 애플리케이션 (354) 은 검출된 사용자 음성에 기초하여 하나 이상의 동작을 수행하도록 구성될 수 있다. 예시를 위해, 애플리케이션 (354) 은, 예시적이고 비제한적인 예들로서, 음성 인터페이스 애플리케이션, 통합 어시스턴트 애플리케이션, 비히클 내비게이션 및 엔터테인먼트 애플리케이션, 또는 홈 자동화 시스템에 대응할 수 있다.

[0063] 자가 음성 활동 검출기 (320) 의 제 1 스테이지 (340) 에서 오디오 데이터를 프로세싱한 결과에 기초하여 제 2 스테이지 (350) 를 선택적으로 활성화함으로써, 자가 음성 활동 검출, 음성 커맨드 프로세싱, 또는 양자 모두와 연관된 전체 전력 소비가 감소될 수도 있다.

[0064] 도 4 는 본 개시의 일부 예들에 따른, 도 1 의 시스템의 컴포넌트들의 동작의 예시적인 양태의 다이어그램이다.

특징 추출기 (130) 는 제 1 프레임 (F1)(412), 제 2 프레임 (F2)(414), 및 제 N 프레임 (FN)(416)(여기서 N 은 2보다 큰 정수) 을 포함하는 하나 이상의 추가 프레임들로서 예시된, 오디오 데이터 (128) 의 연속적으로 캡

처된 프레임들의 시퀀스와 같은 오디오 데이터 샘플들의 시퀀스 (410) 를 수신하도록 구성된다. 특징 추출기 (130) 는 제 1 세트 (422), 제 2 세트 (424), 및 제 N 세트 (426) 를 포함하는 하나 이상의 추가 세트를 포함하는 특징 데이터 세트들의 시퀀스 (420) 를 출력하도록 구성된다.

[0065] 동적 분류기 (140) 는 특징 데이터의 세트들의 시퀀스 (420) 를 수신하고, 시퀀스 (420) 에서 특징 데이터의 이전 세트 (예를 들어, 제 1 세트 (422)) 에 적어도 부분적으로 기초하여 시퀀스 (420) 의 세트 (예를 들어, 제 2 세트 (424)) 를 적응적으로 클러스터링하도록 구성된다. 예시적이고 비제한적인 예들로서, 동적 분류기 (140) 는 시간 Kohonen 맵 또는 순환 자기-조직화 맵으로서 구현될 수 있다.

[0066] 동작 동안, 특징 추출기 (130) 는 특징 데이터의 제 1 세트 (422) 를 생성하기 위해 제 1 프레임 (412) 을 프로세싱하고, 동적 분류기 (140) 는 분류 출력들의 시퀀스 (430) 의 제 1 분류 출력 (C1)(432) 을 생성하기 위해 특징 데이터의 제 1 세트 (422) 를 프로세싱한다. 특징 추출기 (130) 는 특징 데이터의 제 2 세트 (424) 를 생성하기 위해 제 2 프레임 (414) 을 프로세싱하고, 동적 분류기 (140) 는 특징 데이터의 제 2 세트 (424) 에 기초하여 그리고 특징 데이터의 제 1 세트 (422) 에 적어도 부분적으로 기초하여 제 2 분류 출력 (C2)(434) 을 생성하기 위해 특징 데이터의 제 2 세트 (424) 를 프로세싱한다. 특징 추출기 (130) 가 특징 데이터의 제 N 세트 (426) 를 생성하기 위해 제 N 프레임 (416) 을 프로세싱하는 것을 포함하는, 이러한 프로세싱이 계속되고, 동적 분류기 (140) 는 제 N 분류 출력 (CN)(436) 을 생성하기 위해 특징 데이터의 제 N 세트 (426) 를 프로세싱한다. 제 N 분류 출력 (436) 은 특징 데이터의 제 N 세트 (426) 에 기초하고 시퀀스 (420) 의 특징 데이터의 이전 세트들 중 하나 이상에 적어도 부분적으로 기초한다.

[0067] 특징 데이터의 하나 이상의 이전 세트들에 기초하여 동적으로 분류함으로써, 동적 분류기 (140) 에 의한 분류의 정확도는 오디오 데이터의 다수의 프레임들에 걸쳐 있을 수 있는 스피치 신호들에 대해 개선될 수 있다.

[0068] 도 5 는 하나 이상의 프로세서들 (190) 을 포함하는 집적 회로 (502) 로서의 디바이스 (102) 의 구현 (500) 을 도시한다. 집적 회로 (502) 는 또한 오디오 데이터 (128) 가 프로세싱을 위해 수신될 수 있게 하는 하나 이상의 버스 인터페이스들과 같은 오디오 입력 (504) 을 포함한다. 집적 회로 (502) 는 또한 사용자 음성 활동 표시기 (150) 와 같은 출력 신호의 전송을 가능하게 하기 위해 버스 인터페이스와 같은 신호 출력 (512) 을 포함한다. 집적 회로 (502) 는 도 6 에 도시된 바와 같은 모바일 폰 또는 태블릿, 도 7 에 도시된 바와 같은 헤드셋, 도 8 에 도시된 바와 같은 웨어러블 전자 디바이스, 도 9 에 도시된 바와 같은 음성 제어 스피커 시스템, 도 10 에 도시된 바와 같은 카메라, 도 11 에 도시된 바와 같은 가상 현실 헤드셋, 혼합 현실 헤드셋 또는 증강 현실 헤드셋, 또는 도 12 또는 도 13 에 도시된 바와 같은 비히클과 같은 마이크로폰들을 포함하는 시스템에서의 컴포넌트로서 자기 음성 활동 검출의 구현을 가능하게 한다.

[0069] 도 6 은, 예시적이고 비제한적인 예들로서, 디바이스 (102) 가 모바일 디바이스 (602), 예를 들어 전화 또는 태블릿인 구현 (600) 을 도시한다. 모바일 디바이스 (602) 는 사용자의 스피치를 일차적으로 캡처하도록 위치설정된 제 1 마이크로폰 (110), 환경 사운드들을 일차적으로 캡처하도록 위치설정된 다수의 제 2 마이크로폰들 (120), 및 디스플레이 스크린 (604) 을 포함한다. 특징 추출기 (130) 및 동적 분류기 (140) 를 포함하는 프로세서 (190) 의 컴포넌트들은 모바일 디바이스 (602) 에 통합되고, 모바일 디바이스 (602) 의 사용자에게 일반적으로 보이지 않는 내부 컴포넌트들을 표시하기 위해 점선들을 사용하여 예시된다. 프로세서 (190) 가 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다. 특정 예에서, 동적 분류기 (140) 는 사용자 음성 활동을 검출하도록 동작하고, 이는 그후 그래픽 사용자 인터페이스를 론치하거나 그렇지 않으면 (예를 들어, 통합된 "스마트 어시스턴트" 애플리케이션을 통해) 디스플레이 스크린 (604) 에서 사용자의 스피치와 연관된 다른 정보를 디스플레이하는 것과 같이, 모바일 디바이스 (602) 에서 하나 이상의 동작을 수행하도록 프로세싱된다.

[0070] 도 7 은 디바이스 (102) 가 헤드셋 디바이스 (702) 인 구현 (700) 을 도시한다. 헤드셋 디바이스 (702) 는 사용자의 스피치를 일차적으로 캡처하도록 위치설정된 제 1 마이크로폰 (110) 및 환경 사운드들을 일차적으로 캡처하도록 위치설정된 제 2 마이크로폰 (120) 을 포함한다. 특징 추출기 (130) 및 동적 분류기 (140) 를 포함하는 프로세서 (190) 의 컴포넌트들은 헤드셋 디바이스 (702) 에 통합된다. 특정 예에서, 동적 분류기 (140) 는 사용자 음성 활동을 검출하도록 동작하며, 이는 헤드셋 디바이스 (702) 로 하여금 헤드셋 디바이스 (702) 에서 하나 이상의 동작들을 수행하거나, 추가 프로세싱을 위해, 사용자 음성 활동에 대응하는 오디오 데이터를 도 1 의 제 2 디바이스 (160) 와 같은 제 2 디바이스 (미도시) 에 송신하거나, 또는 그 조합을 가능하게

한다. 프로세서 (190) 가 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다.

[0071] 도 8 은 디바이스 (102) 가 "스마트 워치"로서 예시된 웨어러블 전자 디바이스 (802) 인 구현 (800) 을 도시한다. 특징 추출기 (130), 동적 분류기 (140), 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 은 웨어러블 전자 디바이스 (802) 에 통합된다. 특정 예에서, 동적 분류기 (140) 는 사용자 음성 활동을 검출하도록 동작하고, 이는 그후 그래픽 사용자 인터페이스를 론치하거나 그렇지 않으면 웨어러블 전자 디바이스 (802) 의 디스플레이 스크린 (804) 에서 사용자의 스피치와 연관된 다른 정보를 디스플레이하는 것과 같이 웨어러블 전자 디바이스 (802) 에서 하나 이상의 동작을 수행하도록 프로세싱된다. 예시를 위해, 웨어러블 전자 디바이스 (802) 는 웨어러블 전자 디바이스 (802) 에 의해 검출된 사용자 음성에 기초하여 통지를 디스플레이하도록 구성된 디스플레이 스크린을 포함할 수 있다. 특정 예에서, 웨어러블 전자 디바이스 (802) 는 사용자 음성 활동의 검출에 응답하여 햅틱 통지를 제공하는 (예를 들어, 진동하는) 햅틱 디바이스를 포함한다. 예를 들어, 햅틱 통지는 사용자가 웨어러블 전자 디바이스 (802) 를 바라보게 하여 사용자에 의해 스피치된 키워드의 검출을 표시하는 디스플레이된 통지를 보게 할 수 있다. 따라서, 웨어러블 전자 디바이스 (802) 는 청각 장애를 갖는 사용자 또는 헤드셋을 착용한 사용자에게 사용자의 음성 활동이 검출된 것을 알릴 수 있다. 웨어러블 전자 디바이스 (802) 가 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다.

[0072] 도 9 는 디바이스 (102) 가 무선 스피커 및 음성 활성화 디바이스 (902) 인 구현 (900) 이다. 무선 스피커 및 음성 활성화 디바이스 (902) 는 무선 네트워크 연결성을 가질 수 있고 어시스턴트 동작을 실행하도록 구성된다. 특징 추출기 (130) 및 동적 분류기 (140) 를 포함하는 프로세서 (190), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 또는 그 조합은 무선 스피커 및 음성 활성화 디바이스 (902) 에 포함된다. 프로세서 (190) 가 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다. 무선 스피커 및 음성 활성화 디바이스 (902) 는 또한 스피커 (904) 를 포함한다. 동작 동안, 동적 분류기 (140) 의 동작을 통해 사용자 스피치로서 식별된 구두 커맨드를 수신하는 것에 응답하여, 무선 스피커 및 음성 활성화 디바이스 (902) 는 음성 활성화 시스템 (162)(예를 들어, 통합 어시스턴트 애플리케이션) 의 실행을 통해서와 같이 어시스턴트 동작들을 실행할 수 있다. 어시스턴트 동작들은 온도를 조정하는 것, 음악을 재생하는 것, 조명들을 켜는 것 등을 포함할 수 있다. 예를 들어, 어시스턴트 동작들은 키워드 또는 키 구문 (예를 들어, "헬로 어시스턴트") 이후의 커맨드를 수신하는 것에 응답하여 수행된다.

[0073] 예시적인 예에서, 무선 스피커 및 음성 활성화 디바이스 (902) 가 (예를 들어, 윈도우 옆에) 룸의 벽 근처에 있고, 제 1 마이크로폰 (110) 이 제 2 마이크로폰 (120) 에 비해 룸의 내부에 더 가깝게 배열되도록 배열될 때 (예를 들어, 제 2 마이크로폰은 제 1 마이크로폰 (110) 보다 벽 또는 윈도우에 더 가깝게 위치설정될 수 있음), 룸의 내부로부터 기원하는 스피치는 사용자 음성 활동으로서 식별될 수 있는 반면, 룸 외부로부터 기원하는 사운드 (예를 들어, 벽 또는 윈도우의 다른 측 상의 사람의 스피치) 는 다른 오디오 활동으로서 식별될 수 있다.

다수의 사람들이 룸에 있을 수도 있기 때문에, 무선 스피커 및 음성 활성화 디바이스 (902) 는 다수의 사람들 중 임의의 사람으로부터의 스피치를 사용자 음성 활동으로서 식별하도록 구성될 수도 있다 (예를 들어, 무선 스피커 및 음성 활성화 디바이스 (902) 의 다수의 "사용자들" 이 존재할 수도 있다). 예시를 위해, 동적 분류기 (140) 는 말하는 사람이 무선 스피커 및 음성 활성화 디바이스 (902) 로부터 비교적 멀리 (예를 들어, 수 미터) 있을 수도 있고 제 2 마이크로폰 (120) 보다 제 1 마이크로폰 (110) 에 더 가까울 때에도 룸 내로부터 기인하는 스피치에 대응하는 특징 데이터를 "자기-음성" 으로서 인식하도록 구성될 수도 있다. 스피치가 룸에서 다수의 사람들로부터 검출되는 일부 구현들에서, 무선 스피커 및 음성 활성화 디바이스 (902)(예를 들어, 동적 분류기 (140)) 는 제 1 마이크로폰 (110) 에 가장 가까운 사람으로부터의 스피치를 사용자 음성 활동 (예를 들어, 가장 가까운 사용자와의 자기-음성) 으로서 식별하도록 구성될 수도 있다.

[0074] 도 10 은 디바이스 (102) 가 카메라 디바이스 (1002) 에 대응하는 휴대용 전자 디바이스인 구현 (1000) 을 도시한다. 특징 추출기 (130) 및 동적 분류기 (140), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 또는 그 조합은 카메라 디바이스 (1002) 에 포함된다. 동작 동안, 동적 분류기 (140) 의 동작을 통해 사용자 스피치로서 식별된 구두 커맨드를 수신하는 것에 응답하여, 카메라 디바이스 (1002) 는 예시적인 예들로서, 이미지 또

는 비디오 캡처 설정들, 이미지 또는 비디오 플레이백 설정들, 또는 이미지 또는 비디오 캡처 명령들을 조정하는 것과 같이, 스피치된 사용자 커맨드들에 응답하여 동작들을 실행할 수 있다. 카메라 (1002) 가 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다.

[0075] 도 11 은 디바이스 (102) 가 가상 현실 ("VR"), 증강 현실 ("AR"), 또는 혼합 현실 ("MR") 헤드셋 디바이스와 같은, 확장 현실 ("XR") 헤드셋 (1102) 에 대응하는 휴대용 전자 디바이스를 포함하는 구현 (1100) 을 도시한다. 특징 추출기 (130), 동적 분류기 (140), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 또는 그 조합은 헤드셋 (1102) 에 통합된다. 특정 양태에서, 헤드셋 (1102) 은 사용자의 스피치를 일차적으로 캡처하도록 위치설정된 제 1 마이크로폰 (110) 및 환경 사운드들을 일차적으로 캡처하도록 위치설정된 제 2 마이크로폰 (120) 을 포함한다. 사용자 음성 활동 검출은 헤드셋 (1102) 의 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 으로부터 수신된 오디오 신호들에 기초하여 수행될 수 있다. 시각적 인터페이스 디바이스는 헤드셋 (1102) 이 착용되는 동안 사용자로의 증강 현실 또는 가상 현실 이미지들 또는 장면들의 디스플레이를 가능하게 하기 위해 사용자의 눈 앞에 위치설정된다. 특정 예에서, 시각적 인터페이스 디바이스는 오디오 신호에서 검출된 사용자 스피치를 표시하는 통지를 디스플레이하도록 구성된다. 헤드셋 (1102) 이 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다.

[0076] 도 12 는 디바이스 (102) 가 유인 또는 무인 항공기 디바이스 (예를 들어, 패키지 배송 드론) 로서 예시된 비히클 (1202) 에 대응하거나 그 안에 통합되는 구현 (1200) 을 도시한다. 특징 추출기 (130), 동적 분류기 (140), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 또는 그 조합은 비히클 (1202) 에 통합된다. 사용자 음성 활동 검출은, 예를 들어 비히클 (1202) 의 허가된 사용자로부터의 명령들의 전달을 위해서, 비히클 (1202) 의 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 으로부터 수신된 오디오 신호들에 기초하여 수행될 수 있다. 비히클 (1202) 이 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다.

[0077] 도 13 은 디바이스 (102) 가 자동차로서 예시된 비히클 (1302) 에 대응하거나 그 안에 통합되는 또 다른 구현 (1300) 을 도시한다. 비히클 (1302) 은 특징 추출기 (130) 및 동적 분류기 (140) 를 포함하는 프로세서 (190) 를 포함한다. 비히클 (1302) 이 특징 추출기 (130) 를 포함하는 것으로 예시되지만, 다른 구현들에서, 예를 들어 동적 분류기 (140) 가 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터를 추출하도록 구성될 때, 특징 추출기 (130) 는 생략된다. 비히클 (1302) 은 또한 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 을 포함한다. 제 1 마이크로폰 (110) 은 비히클 (1302) 의 조작자의 발화들을 캡처하도록 위치설정된다. 사용자 음성 활동 검출은 비히클 (1302) 의 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 으로부터 수신된 오디오 신호들에 기초하여 수행될 수 있다. 일부 구현들에서, 사용자 음성 활동 검출은, 예를 들어 허가된 승객으로부터의 음성 커맨드에 대해, 내부 마이크로폰들 (예를 들어, 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120)) 로부터 수신된 오디오 신호에 기초하여 수행될 수 있다. 예를 들어, 사용자 음성 활동 검출은 비히클 (1302) 의 조작자로부터의 (예를 들어, 볼륨을 5 로 설정하거나 자율 주행 비히클에 대한 목적지를 설정하기 위한 부모로부터의) 음성 커맨드를 검출하고 또 다른 승객의 음성 (예를 들어, 볼륨을 10 으로 설정하기 위한 어린이로부터의 또는 또 다른 위치를 논의하는 다른 승객들로부터의 커맨드) 을 무시하는 데 사용될 수 있다. 일부 구현들에서, 사용자 음성 활동 검출은 비히클의 허가된 사용자와 같은 외부 마이크로폰들 (예를 들어, 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120)) 로부터 수신된 오디오 신호에 기초하여 수행될 수 있다. 특정 구현에서, 동적 분류기 (140) 의 동작을 통해 사용자 스피치로서 식별된 구두 커맨드를 수신하는 것에 응답하여, 음성 활성화 시스템 (162) 은, 예를 들어, 디스플레이 (1320) 또는 하나 이상의 스피커 (예를 들어, 스피커 (1310)) 를 통해 피드백 또는 정보를 제공함으로써, 출력 신호 (135) 에서 검출된 하나 이상의 키워드 (예를 들어, "잠금 해제", "엔진 시작", "음악 플레이", "일기 예보 디스플레이" 또는 또 다른 음성 명령) 에 기초하여 비히클 (1302) 의 하나 이상의 동작을 개시한다.

[0078] 도 14a 를 참조하면, 사용자 음성 활동 검출의 방법 (1400) 의 특정 구현이 도시된다. 특정 양태에서, 방법 (1400) 의 하나 이상의 동작들은 도 1 의 특징 추출기 (130), 동적 분류기 (140), 프로세서 (190), 디바이스

(102), 시스템 (100), 또는 그 조합 중 적어도 하나에 의해 수행된다.

[0079] 방법 (1400) 은 1402 에서, 하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계를 포함한다. 예를 들어, 도 1 의 특징 추출기 (130) 는 도 1 을 참조하여 설명된 바와 같이, 제 1 마이크로폰 (110) 의 제 1 출력에 대응하는 제 1 오디오 데이터 (116) 및 제 2 마이크로폰 (126) 의 제 2 출력에 대응하는 제 2 오디오 데이터 (126) 를 포함하는 오디오 데이터 (128) 를 수신한다.

[0080] 방법 (1400) 은 1404 에서, 하나 이상의 프로세서들에서, 제 1 오디오 데이터 및 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하는 단계를 포함한다. 예를 들어, 도 1 의 특징 추출기 (130) 는 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 에 기초하여 특징 데이터 (132) 를 생성한다. 또 다른 예에서, 도 1 의 동적 분류기 (140) 와 같은 동적 분류기는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 수신하고 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 의 프로세싱 동안 특징 데이터 (132) 를 추출하도록 구성된다.

[0081] 방법 (1400) 은 1406 에서, 하나 이상의 프로세서들의 동적 분류기에서, 특징 데이터의 분류 출력을 생성하는 단계를 포함한다. 예를 들어, 도 1 의 동적 분류기 (140) 는 도 1 을 참조하여 설명된 바와 같이 특징 데이터 (132) 의 분류 출력 (142) 을 생성한다.

[0082] 방법 (1400) 은 1408 에서, 하나 이상의 프로세서들에서 그리고 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함한다. 예를 들어, 도 1 의 프로세서 (190) 는 도 1 을 참조하여 설명된 바와 같이, 분류 출력 (142) 에 적어도 부분적으로 기초하여, 오디오 데이터 (128) 가 사용자 음성 활동에 대응하는지 여부를 결정한다.

[0083] 방법 (1400) 은 종래의 자기-음성 활동 검출 기술들에 비해 상대적으로 낮은 복잡도, 낮은 전력 소비, 및 높은 정확도로 사용자 음성 활동과 다른 오디오 활동 사이에서 구별하기 위해 동적 분류기 (140) 를 사용함으로써 자기-음성 활동 검출의 성능을 개선한다. 사용자 및 환경 변화에 자동으로 적응하는 것은 사용자에 의해 수행되는 캘리브레이션을 감소시키거나 제거함으로써 개선된 이점을 제공하고 사용자의 경험을 향상시킨다.

[0084] 도 14b 를 참조하면, 사용자 음성 활동 검출의 방법 (1450) 의 특정 구현이 도시된다. 특정 양태에서, 방법 (1450) 의 하나 이상의 동작들은 도 1 의 동적 분류기 (140), 프로세서 (190), 디바이스 (102), 시스템 (100), 또는 그 조합 중 적어도 하나에 의해 수행된다.

[0085] 방법 (1450) 은 1452 에서, 하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계를 포함한다. 예에서, 도 1 의 특징 추출기 (130) 는 도 1 을 참조하여 설명된 바와 같이, 제 1 오디오 데이터 (116) 및 제 2 마이크로폰 (126) 의 제 2 출력에 대응하는 제 2 오디오 데이터 (126) 를 포함하는 오디오 데이터 (128) 를 수신한다.

[0086] 방법 (1450) 은 1454 에서, 하나 이상의 프로세서들에서, 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 오디오 데이터를 동적 분류기에 제공하는 단계를 포함한다. 일 예에서, 도 1 의 특징 추출기 (130) 는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 에 기초하여 특징 데이터 (132) 를 생성하고, 특징 데이터 (132) 는 예를 들어 도 1 에 설명된 바와 같이 그리고 도 14a 의 방법 (1400) 에 따라 분류 출력 (142) 을 생성하기 위해 동적 분류기 (140) 에 의해 프로세싱된다. 또 다른 예에서, 프로세서 (190) 는 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 동적 분류기 (140) 에 제공하고, 동적 분류기 (140) 는 분류 출력 (142) 을 생성하기 위해 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 프로세싱한다. 예시적인 구현에서, 동적 분류기 (140) 는 특징 데이터 (132) 를 추출하기 위해 제 1 오디오 데이터 (116) 및 제 2 오디오 데이터 (126) 를 프로세싱하고, 특징 데이터 (132) 에 기초하여 분류 출력 (142) 을 결정한다.

[0087] 방법 (1450) 은 1456 에서, 하나 이상의 프로세서들에서 그리고 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함한다. 예를 들어, 도 1 의 프로세서 (190) 는 도 1 을 참조하여 설명된 바와 같이, 분류 출력 (142) 에 적어도 부분적으로 기초하여, 오디오 데이터 (128) 가 사용자 음성 활동에 대응하는지 여부를 결정한다.

[0088] 방법 (1450) 은 종래의 자기-음성 활동 검출 기술들에 비해 상대적으로 낮은 복잡도, 낮은 전력 소비, 및 높은 정확도로 사용자 음성 활동과 다른 오디오 활동 사이에서 구별하기 위해 동적 분류기 (140) 를 사용함으로써 자기-음성 활동 검출의 성능을 개선한다. 사용자 및 환경 변화에 자동으로 적응하는 것은 사용자에 의해 수행

되는 캘리브레이션을 감소시키거나 제거함으로써 개선된 이점을 제공하고 사용자의 경험을 향상시킨다.

[0089] 도 14a 의 방법 (1400), 도 14b 의 방법 (1450), 또는 그 조합은, 필드 프로그래밍가능 게이트 어레이 (FPGA) 디바이스, 주문형 집적 회로 (ASIC), 프로세싱 유닛, 예를 들어 중앙 프로세싱 유닛 (CPU), DSP, 제어기, 다른 하드웨어 디바이스, 펌웨어 디바이스, 또는 이들의 임의의 조합에 의해 구현될 수도 있다. 일 예로서, 도 14a 의 방법 (1400), 도 14b 의 방법 (1450), 또는 그 조합은 도 15 를 참조하여 설명된 바와 같은 명령들을 실행하는 프로세서에 의해 수행될 수도 있다.

[0090] 도 15 를 참조하면, 디바이스의 특정 예시적인 구현의 블록 다이어그램이 도시되고 일반적으로 1500 으로 지정된다. 다양한 구현들에서, 디바이스 (1500) 는 도 15 에 예시된 것보다 더 많거나 또는 더 적은 컴포넌트들을 가질 수도 있다. 예시적인 구현에서, 디바이스 (1500) 는 디바이스 (102) 에 대응할 수도 있다. 예시적인 구현에서, 디바이스 (1500) 는 도 1 내지 도 14b 를 참조하여 설명된 하나 이상의 동작들을 수행할 수도 있다.

[0091] 특정 구현에서, 디바이스 (1500) 는 프로세서 (1506)(예를 들어, 중앙 프로세싱 유닛 (CPU)) 를 포함한다. 디바이스 (1500) 는 하나 이상의 추가적인 프로세서들 (1510)(예를 들어, 하나 이상의 DSP들) 을 포함할 수도 있다. 특정 양태에서, 도 1 의 프로세서 (190) 는 프로세서 (1506), 프로세서들 (1510), 또는 그 조합에 대응한다. 프로세서들 (1510) 은 음성 코더 ("보코더") 인코더 (1536), 보코더 디코더 (1538), 특징 추출기 (130), 동적 분류기 (140), 또는 그 조합을 포함하는 스피치 및 음악 코더-디코더 (CODEC)(1508) 를 포함할 수도 있다.

[0092] 디바이스 (1500) 는 메모리 (1586) 및 CODEC (1534) 을 포함할 수도 있다. 메모리 (1586) 는 특징 추출기 (130), 동적 분류기 (140), 또는 양자 모두를 참조하여 설명된 기능을 구현하기 위해 하나 이상의 추가적인 프로세서들 (1510)(또는 프로세서 (1506)) 에 의해 실행가능한 명령들 (1556) 을 포함할 수도 있다. 디바이스 (1500) 는, 트랜시버 (1550) 를 통해, 안테나 (1552) 에 커플링된 모뎀 (170) 을 포함할 수도 있다.

[0093] 디바이스 (1500) 는 디스플레이 제어기 (1526) 에 커플링된 디스플레이 (1528) 를 포함할 수도 있다. 스피커 (1592), 제 1 마이크로폰 (110), 및 제 2 마이크로폰 (120) 은 CODEC (1534) 에 커플링될 수도 있다. CODEC (1534) 은 디지털-아날로그 변환기 (DAC)(1502), 아날로그-디지털 변환기 (ADC)(1504) 또는 양자 모두를 포함할 수도 있다. 특정 구현에서, 코덱 (1534) 은 제 1 마이크로폰 (110) 및 제 2 마이크로폰 (120) 으로부터 아날로그 신호들을 수신하고, 아날로그-디지털 변환기 (1504) 를 사용하여 아날로그 신호들을 디지털 신호들로 변환하고, 디지털 신호들을 스피치 및 음악 코덱 (1508) 에 제공할 수도 있다. 스피치 및 음악 코덱 (1508) 은 디지털 신호들을 프로세싱할 수도 있고, 디지털 신호들은 특징 추출기 (130) 및 동적 분류기 (140) 에 의해 추가로 프로세싱될 수도 있다. 특정 구현에서, 스피치 및 음악 코덱 (1508) 은 CODEC (1534) 에 디지털 신호들을 제공할 수도 있다. 코덱 (1534) 은 디지털 신호들을 디지털-아날로그 변환기 (1502) 를 사용하여 아날로그 신호들로 변환할 수도 있고 아날로그 신호들을 스피커 (1592) 에 제공할 수도 있다.

[0094] 특정 구현에서, 디바이스 (1500) 는 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1522) 에 포함될 수도 있다. 특정 구현에서, 메모리 (1586), 프로세서 (1506), 프로세서들 (1510), 디스플레이 제어기 (1526), 코덱 (1534), 및 모뎀 (170) 은 시스템-인-패키지 또는 시스템-온-칩 디바이스 (1522) 에 포함된다. 특정의 구현예에서, 입력 디바이스 (1530) 및 전력 공급부 (1544) 는 시스템-온-칩 디바이스 (1522) 에 커플링된다. 또한, 특정 구현에서, 도 15 에 예시된 바와 같이, 디스플레이 (1528), 입력 디바이스 (1530), 스피커 (1592), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 안테나 (1552), 및 전력 공급부 (1544) 는 시스템-온-칩 디바이스 (1522) 외부에 있다. 특정 구현에서, 디스플레이 (1528), 입력 디바이스 (1530), 스피커 (1592), 제 1 마이크로폰 (110), 제 2 마이크로폰 (120), 안테나 (1552), 및 전력 공급부 (1544) 의 각각은 인터페이스 (예를 들어, 제 1 입력 인터페이스 (114) 또는 제 2 입력 인터페이스 (124)) 또는 제어기와 같은 시스템-온-칩 디바이스 (1522) 의 컴포넌트에 커플링될 수도 있다.

[0095] 디바이스 (1500) 는 스마트 스피커, 스피커 바, 모바일 통신 디바이스, 스마트 폰, 셀룰러 폰, 랩탑 컴퓨터, 컴퓨터, 태블릿, 개인용 디지털 보조기, 디스플레이 디바이스, 텔레비전, 게이밍 콘솔, 뮤직 플레이어, 라디오, 디지털 비디오 플레이어, 디지털 비디오 디스크 (DVD) 플레이어, 튜너, 카메라, 내비게이션 디바이스, 비히클, 헤드셋, 증강 혼합 현실 헤드셋, 가상 현실 헤드셋, 항공 비히클, 홈 자동화 시스템, 음성 활성화 디바이스, 무선 스피커 및 음성 활성화 디바이스, 휴대용 전자 디바이스, 자동차, 비히클, 컴퓨팅 디바이스, 통신 디바이스, 사물 인터넷 (IoT) 디바이스, 가상 현실 (VR) 디바이스, 기지국, 모바일 디바이스, 또는 이들의 임의의 조합을 포함할 수도 있다.

[0096] 설명된 구현과 함께, 장치는 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 수신하기 위한 수단을 포함한다. 예를 들어, 수신하기 위한 수단은 제 1 입력 인터페이스 (114), 제 2 입력 인터페이스 (124), 특징 추출기 (130), 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다.

[0097] 장치는 또한 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하기 위한 수단을 포함한다. 예를 들어, 특징 데이터를 생성하기 위한 수단은 특징 추출기 (130), 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 특징 데이터를 생성하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다.

[0098] 장치는 동적 분류기에서, 특징 데이터의 분류 출력을 생성하기 위한 수단을 더 포함한다. 예를 들어, 분류 출력을 생성하기 위한 수단은 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 동적 분류기에서 분류 출력을 생성하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다.

[0099] 장치는 또한 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단을 포함한다. 예를 들어, 결정하기 위한 수단은 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다

[0100] 설명된 구현과 함께, 장치는 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 수신하기 위한 수단을 포함한다. 예를 들어, 수신하기 위한 수단은 제 1 입력 인터페이스 (114), 제 2 입력 인터페이스 (124), 특징 추출기 (130), 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다.

[0101] 장치는 추가로 동적 분류기에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위한 수단를 포함한다. 예를 들어, 분류 출력을 생성하기 위한 수단은 특징 추출기 (130), 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 동적 분류기에서 분류 출력을 생성하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다.

[0102] 장치는 또한 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단을 포함한다. 예를 들어, 결정하기 위한 수단은 동적 분류기 (140), 프로세서 (190), 하나 이상의 프로세서들 (1510), 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성된 하나 이상의 다른 회로들 또는 컴포넌트들, 또는 그 임의의 조합에 대응할 수 있다

[0103] 일부 구현들에서, 비일시적 컴퓨터 판독가능 매체 (예를 들어, 메모리 (1586) 와 같은 컴퓨터 판독가능 저장 디바이스) 는, 하나 이상의 프로세서들 (예를 들어, 하나 이상의 프로세서들 (1510) 또는 프로세서 (1506)) 에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 제 1 마이크로폰 (예를 들어, 제 1 마이크로폰 (110)) 의 제 1 출력에 대응하는 제 1 오디오 데이터 (예를 들어, 제 1 오디오 데이터 (116)) 및 제 2 마이크로폰 (예를 들어, 제 2 마이크로폰 (120)) 의 제 2 출력에 대응하는 제 2 오디오 데이터 (예를 들어, 제 2 오디오 데이터 (126)) 를 포함하는 오디오 데이터 (예를 들어, 오디오 데이터 (128)) 를 수신하게 하는 명령들 (예를 들어, 명령들 (1556)) 을 포함한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 또한, 하나 이상의 프로세서들로 하여금, 오디오 데이터에 대응하는 분류 출력 (예를 들어, 분류 출력 (142)) 을 생성하기 위해 오디오 데이터를 동적 분류기 (예를 들어, 동적 분류기 (140)) 에 제공하게 한다. 일 예에서, 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 하나 이상의 프로세서들로 하여금, 제 1 오디오 데이터 및 제 2 오디오 데이터에 기초하여 특징 데이터 (예를 들어, 특징 데이터 (132)) 를 생성하게 하고 동적 분류기에서 특징 데이터를 프로세싱하게 한다. 명령들은, 하나 이상의 프로세서들에 의해 실행될 때, 또한, 하나 이상의 프로세서들로 하여금, 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결

정하게 한다.

[0104] 본 개시는 다음의 예들을 포함한다.

[0105] 예 1. 하나 이상의 프로세서들을 포함하는 디바이스로서, 상기 하나 이상의 프로세서들은, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고; 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하고; 상기 특징 데이터의 분류 출력을 생성하기 위해 동적 분류기에서 상기 특징 데이터를 프로세싱하고; 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성된다.

[0106] 예 2. 예 1 의 디바이스에 있어서, 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 사용자의 발화들을 캡처하도록 구성되고, 상기 제 2 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 주변 사운드를 캡처하도록 구성된다.

[0107] 예 3. 예 1 의 디바이스에 있어서, 상기 특징 데이터는, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함한다.

[0108] 예 4. 예 3 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하도록 추가로 구성되고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함한다.

[0109] 예 5. 예 1 의 디바이스에 있어서, 상기 동적 분류기는 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 적응적으로 클러스터링하도록 구성된다.

[0110] 예 6. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 특징 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하도록 추가로 구성된다.

[0111] 예 7. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 동적 분류기의 분류 결정 기준을 업데이트하도록 추가로 구성된다.

[0112] 예 8. 예 1 의 디바이스에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0113] 예 9. 예 1 의 디바이스에 있어서, 상기 동적 분류기는 특징 데이터의 세트들의 시퀀스를 수신하고, 상기 시퀀스에서의 특징 데이터의 이전 세트에 적어도 부분적으로 기초하여 상기 시퀀스의 세트를 적응적으로 클러스터링하도록 추가로 구성된다.

[0114] 예 10. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초하여 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성된다.

[0115] 예 11. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하도록 추가로 구성된다.

[0116] 예 12. 예 11 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하도록 구성된다.

[0117] 예 13. 예 12 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은, 추가로, 상기 동적 분류기를 포함하는 올웨이즈온 (always-on) 전력 도메인; 및 음성 커맨드 프로세싱 유닛을 포함하는 제 2 전력 도메인을 포함하고, 상기 웨이크업 신호는 상기 음성 커맨드 프로세싱 유닛을 활성화하기 위해 저전력 모드로부터 상기 제 2 전력 도메인을 전이하도록 구성된다.

[0118] 예 14. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들에 커플링된 모뎀을 추가로 포함하고, 상기 모뎀은, 상기 오디오 데이터가 상기 동적 분류기에 기초한 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하도록 구성된다.

[0119] 예 15. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2

마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기 위해 사용자의 입에 상기 제 2 마이크로폰보다 가깝게 상기 제 1 마이크로폰을 위치설정하도록 구성된다.

[0120] 예 16. 예 1 의 디바이스에 있어서, 하나 이상의 프로세서들은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 또는 증강 현실 헤드셋 중 적어도 하나에 통합된다.

[0121] 예 17. 예 1 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 비히클에 통합되고, 상기 비히클은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 비히클의 조작자의 발화들을 캡처하도록 위치설정된다.

[0122] 예 18. 음성 활동 검출 방법으로서, 하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계; 상기 하나 이상의 프로세서들에서, 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하는 단계; 상기 하나 이상의 프로세서들의 동적 분류기에서, 특징 데이터의 분류 출력을 생성하는 단계; 및 하나 이상의 프로세서들에서 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함한다.

[0123] 예 19. 예 18 의 방법에 있어서, 상기 제 1 마이크로폰은 사용자의 발화들을 캡처하도록 구성되고, 상기 제 2 마이크로폰은 주변 사운드를 캡처하도록 구성된다.

[0124] 예 20. 예 18 의 방법에 있어서, 상기 특징 데이터는: 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함한다.

[0125] 예 21. 예 20 의 방법에 있어서, 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하는 단계를 추가로 포함하고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함한다.

[0126] 예 22. 예 18 의 방법에 있어서, 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 상기 동적 분류기에 의해 적응적으로 클러스터링하는 단계를 추가로 포함한다.

[0127] 예 23. 예 18 의 방법에 있어서, 상기 특징 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하는 단계를 추가로 포함한다.

[0128] 예 24. 예 18 의 방법에 있어서, 상기 동적 분류기의 분류 결정 기준을 업데이트하는 단계를 추가로 포함한다.

[0129] 예 25. 예 18 의 방법에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0130] 예 26. 예 18 의 방법에 있어서, 특징 데이터의 세트들의 시퀀스를 상기 동적 분류기에서 수신하는 단계, 및 상기 시퀀스에서의 특징 데이터의 이전 세트에 적어도 부분적으로 기초하여 상기 시퀀스의 세트를 적응적으로 클러스터링하는 단계를 추가로 포함한다.

[0131] 예 27. 예 18 의 방법에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응하는지 여부를 결정하는 단계는 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초한다.

[0132] 예 28. 예 18 의 방법에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하는 단계를 추가로 포함한다.

[0133] 예 29. 예 28 의 방법에 있어서, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하는 단계를 추가로 포함한다.

[0134] 예 30. 예 29 의 방법에 있어서, 상기 웨이크업 신호는 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 전력 도메인을 저전력 모드로부터 전이하도록 구성된다.

[0135] 예 31. 예 18 의 방법에 있어서, 상기 오디오 데이터가 상기 동적 분류기에 기초하여 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하는 단계를 추가로 포함한다.

[0136] 예 32. 예 18 의 방법에 있어서, 상기 하나 이상의 프로세서들은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2 마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기

위해 사용자의 입에 상기 제 2 마이크로폰보다 가깝게 상기 제 1 마이크로폰을 위치설정한다.

[0137] 예 33. 예 18 의 방법에 있어서, 하나 이상의 프로세서들은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 증강 현실 헤드셋 또는 비히클 중 적어도 하나에 통합된다.

[0138] 예 34. 명령들을 포함하는 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 상기 하나 이상의 프로세서들로 하여금 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고; 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하고; 상기 특징 데이터의 분류 출력을 생성하기 위해 동적 분류기에서 상기 특징 데이터를 프로세싱하고; 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하게 한다.

[0139] 예 35. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 제 1 마이크로폰은 사용자의 발화들을 캡처하도록 구성되고, 상기 제 2 마이크로폰은 주변 사운드를 캡처하도록 구성된다.

[0140] 예 36. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 특징 데이터는 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함한다.

[0141] 예 37. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 상기 하나 이상의 프로세서들로 하여금, 추가로 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하게 하고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함한다.

[0142] 예 38. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 상기 동적 분류기에 의해 적응적으로 클러스터링하게 한다.

[0143] 예 39. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로 상기 특징 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하게 한다.

[0144] 예 40. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로, 동적 분류기의 분류 결정 기준을 업데이트하게 한다.

[0145] 예 41. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0146] 예 42. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로, 특징 데이터의 세트들의 시퀀스를 상기 동적 분류기에 의해 수신하고, 그리고 상기 시퀀스에서의 특징 데이터의 이전 세트에 적어도 부분적으로 기초하여 상기 시퀀스의 세트를 적응적으로 클러스터링하게 한다.

[0147] 예 43. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응하는지 여부를 결정하는 단계는 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초한다.

[0148] 예 44. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하게 한다.

[0149] 예 45. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하게 한다.

[0150] 예 46. 예 45 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 웨이크업 신호는 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 전력 도메인을 저전력 모드로부터 전이하도록 구성된다.

[0151] 예 47. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 명령들은 상기 하나 이상의 프로세서들에 의해 실행될 때, 상기 하나 이상의 프로세서들로 하여금, 추가로 상기 오디오 데이터가 상기 동적 분류기에 기초하여 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하게 한다.

[0152] 예 48. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 하나 이상의 프로세서들은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2 마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기 위해 사용자의 입에 상기 제 2 마이크로폰보다 가깝게 상기 제 1 마이크로폰을 위치설정한다.

[0153] 예 49. 예 34 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 하나 이상의 프로세서들은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 증강 현실 헤드셋 또는 비히클 중 적어도 하나에 통합된다.

[0154] 예 50. 장치로서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하기 위한 수단; 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하기 위한 수단; 동적 분류기에서, 특징 데이터의 분류 출력을 생성하기 위한 수단; 및 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단을 포함한다.

[0155] 예 51. 예 50 의 장치에 있어서, 상기 특징 데이터는: 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함한다.

[0156] 예 52. 예 50 의 장치에 있어서, 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하기 위한 수단을 추가로 포함하고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함한다.

[0157] 예 53. 예 50 의 장치에 있어서, 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 적응적으로 클러스터링하기 위한 수단을 추가로 포함한다.

[0158] 예 54. 예 50 의 장치에 있어서, 상기 특징 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하기 위한 수단을 추가로 포함한다.

[0159] 예 55. 예 50 의 장치에 있어서, 상기 동적 분류기의 분류 결정 기준을 업데이트하기 위한 수단을 추가로 포함한다.

[0160] 예 56. 예 50 의 장치에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0161] 예 57. 예 50 의 장치에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하기 위한 수단을 추가로 포함한다.

[0162] 예 58. 예 50 의 장치에 있어서, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하기 위한 수단을 추가로 포함한다.

[0163] 예 59. 예 50 의 장치에 있어서, 상기 오디오 데이터가 상기 동적 분류기에 기초하여 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하기 위한 수단을 추가로 포함한다.

[0164] 예 60. 예 50 의 장치에 있어서, 상기 오디오 데이터를 수신하기 위한 수단, 상기 특징 데이터를 생성하기 위한 수단, 상기 분류 출력을 생성하기 위한 수단, 및 상기 오디오 데이터가 상기 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2 마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기 위해 상기 사용자의 입에 상기 제 2 마이크로폰보다 더 가깝게 상기 제 1 마이크로폰을 위치설정한다.

[0165] 예 61. 예 50 의 장치에 있어서, 상기 오디오 데이터를 수신하기 위한 수단, 상기 특징 데이터를 생성하기 위한 수단, 상기 분류 출력을 생성하기 위한 수단, 및 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부

를 결정하기 위한 수단은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 증강 현실 헤드셋, 또는 비히클 중 적어도 하나에 통합된다.

[0166] 예 62. 디바이스로서, 명령들을 저장하도록 구성된 메모리 및 하나 이상의 프로세서들을 포함하고, 상기 하나 이상의 프로세서들은, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고; 상기 오디오 데이터에 대응하는 분류 출력을 생성하도록 구성된, 동적 분류기에 상기 오디오 데이터를 제공하고; 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 명령들을 실행하도록 구성된다.

[0167] 예 63. 예 62 의 디바이스에 있어서, 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 사용자의 발화들을 캡처하도록 구성되고, 상기 제 2 마이크로폰은 상기 하나 이상의 프로세서들에 커플링되고 주변 사운드를 캡처하도록 구성된다.

[0168] 예 64. 예 62 또는 예 63 의 디바이스에 있어서, 상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초한다.

[0169] 예 65. 예 62 내지 예 64 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은, 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터에 기초하여 특징 데이터를 생성하고; 그리고 상기 특징 데이터를 상기 동적 분류기에 제공하도록 추가로 구성되고, 상기 분류 출력은 상기 특징 데이터에 기초한다.

[0170] 예 66. 예 65 의 디바이스에 있어서, 상기 특징 데이터는: 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 위상 차이; 및 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 적어도 하나의 귀간 강도 차이를 포함한다.

[0171] 예 67. 예 65 또는 예 66 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초하여 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하도록 구성된다.

[0172] 예 68. 예 65 내지 예 67 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 특징 데이터를 생성하기 전에 상기 제 1 오디오 데이터 및 상기 제 2 오디오 데이터를 변환 도메인으로 변환하도록 추가로 구성되고, 상기 특징 데이터는 다수의 주파수에 대한 귀간 위상 차이 및 다수의 주파수에 대한 귀간 강도 차이를 포함한다.

[0173] 예 69. 예 65 내지 예 68 중 어느 한 예의 디바이스에 있어서, 상기 동적 분류기는 상기 오디오 데이터에 표현된 사운드가 상기 제 2 마이크로폰보다 상기 제 1 마이크로폰에 더 가까운 소스로부터 기원하는지 여부에 기초하여 특징 데이터의 세트들을 적응적으로 클러스터링하도록 구성된다.

[0174] 예 70. 예 62 내지 예 69 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 오디오 데이터에 기초하여 상기 동적 분류기의 클러스터링 동작을 업데이트하도록 추가로 구성된다.

[0175] 예 71. 예 62 내지 예 70 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 동적 분류기의 분류 결정 기준을 업데이트하도록 추가로 구성된다.

[0176] 예 72. 예 62 내지 예 71 중 어느 한 예의 디바이스에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0177] 예 73. 예 62 내지 예 72 중 어느 한 예의 디바이스에 있어서, 상기 동적 분류기는 오디오 데이터의 세트들의 시퀀스를 수신하고, 상기 시퀀스에서의 오디오 데이터의 이전 세트에 적어도 부분적으로 기초하여 상기 시퀀스의 세트를 적응적으로 클러스터링하도록 추가로 구성된다.

[0178] 예 74. 예 62 내지 예 73 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하도록 추가로 구성된다.

[0179] 예 75. 예 74 의 디바이스에 있어서, 상기 하나 이상의 프로세서들은, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하도록 구성된다.

[0180] 예 76. 제 75 항에 있어서, 상기 하나 이상의 프로세서들은, 상기 동적 분류기를 포함하는 올웨이즈온(always-on) 전력 도메인; 및 음성 커맨드 프로세싱 유닛을 포함하는 제 2 전력 도메인을 추가로 포함하고, 상기 웨이크업 신호는 상기 음성 커맨드 프로세싱 유닛을 활성화하기 위해 상기 제 2 전력 도메인을 저전력 모드로부터 전이하도록 구성된다.

[0181] 예 77. 예 62 내지 예 76 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들에 커플링된 모뎀을 추가로 포함하고, 상기 모뎀은, 상기 오디오 데이터가 상기 동적 분류기에 기초한 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하도록 구성된다.

[0182] 예 78. 예 62 내지 예 77 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 포함하는 헤드셋 디바이스에 통합되고, 상기 헤드셋 디바이스는, 사용자에 의해 착용될 때, 상기 제 2 마이크로폰에 비해 더 큰 강도 및 더 적은 지연으로 상기 제 1 마이크로폰에서 상기 사용자의 발화들을 캡처하기 위해 사용자의 입에 상기 제 2 마이크로폰보다 가깝게 상기 제 1 마이크로폰을 위치설정하도록 구성된다.

[0183] 예 79. 예 62 내지 예 77 중 어느 한 예의 디바이스에 있어서, 하나 이상의 프로세서들은 모바일 폰, 태블릿 컴퓨터 디바이스, 웨어러블 전자 디바이스, 카메라 디바이스, 가상 현실 헤드셋, 또는 증강 현실 헤드셋 중 적어도 하나에 통합된다.

[0184] 예 80. 예 62 내지 예 77 중 어느 한 예의 디바이스에 있어서, 상기 하나 이상의 프로세서들은 비히클에 통합되고, 상기 비히클은 상기 제 1 마이크로폰 및 상기 제 2 마이크로폰을 추가로 포함하고, 상기 제 1 마이크로폰은 상기 비히클의 조작자의 발화들을 캡처하도록 위치설정된다.

[0185] 예 81. 음성 활동 검출 방법으로서, 하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하는 단계; 하나 이상의 프로세서들에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 상기 오디오 데이터를 동적 분류기에 제공하는 단계; 및 하나 이상의 프로세서들에서 그리고 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 단계를 포함한다.

[0186] 예 82. 예 81 의 방법에 있어서, 상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초한다.

[0187] 예 83. 예 81 또는 예 82 의 방법에 있어서, 상기 동적 분류기는 자기 조직화 맵 (self-organizing map) 을 포함한다.

[0188] 예 84. 예 81 내지 예 83 중 어느 한 예의 방법에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응하는지 여부를 결정하는 단계는 상기 오디오 데이터에 대응하는 특징 데이터의 적어도 하나의 값의 부호 또는 크기 중 적어도 하나에 추가로 기초한다.

[0189] 예 85. 예 81 내지 예 84 중 어느 한 예의 방법에 있어서, 상기 오디오 데이터가 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 음성 커맨드 프로세싱 동작을 개시하는 단계를 추가로 포함한다.

[0190] 예 86. 예 85 의 방법에 있어서, 상기 음성 커맨드 프로세싱 동작을 개시하기 위해 웨이크업 신호 또는 인터럽트 중 적어도 하나를 생성하는 단계를 추가로 포함한다.

[0191] 예 87. 예 81 내지 예 86 중 어느 한 예의 방법에 있어서, 상기 오디오 데이터가 상기 동적 분류기에 기초하여 상기 사용자 음성 활동에 대응한다는 결정에 응답하여 상기 오디오 데이터를 제 2 디바이스에 송신하는 단계를 추가로 포함한다.

[0192] 예 88. 명령들을 포함하는 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 명령들은, 하나 이상의 프로세서들에 의해 실행될 경우, 상기 하나 이상의 프로세서들로 하여금: 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하고; 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 동적 분류기에 상기 오디오 데이터를 제공하고; 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하는 명령들을 실행하게 한다.

[0193] 예 89. 예 88 의 비일시적 컴퓨터 판독가능 저장 매체에 있어서, 상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초한다.

[0194] 예 90. 장치로서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신하기 위한 수단; 동적 분류기에서, 상기 오디오 데이터에 대응하는 분류 출력을 생성하기 위한 수단; 및 상기 분류 출력에 적어도 부분적으로 기초하여, 상기 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정하기 위한 수단을 포함한다.

[0195] 예 91. 예 90 의 방법에 있어서, 상기 분류 출력은 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 게인 차이, 상기 제 1 오디오 데이터와 상기 제 2 오디오 데이터 사이의 위상 차이, 또는 그 조합에 기초한다.

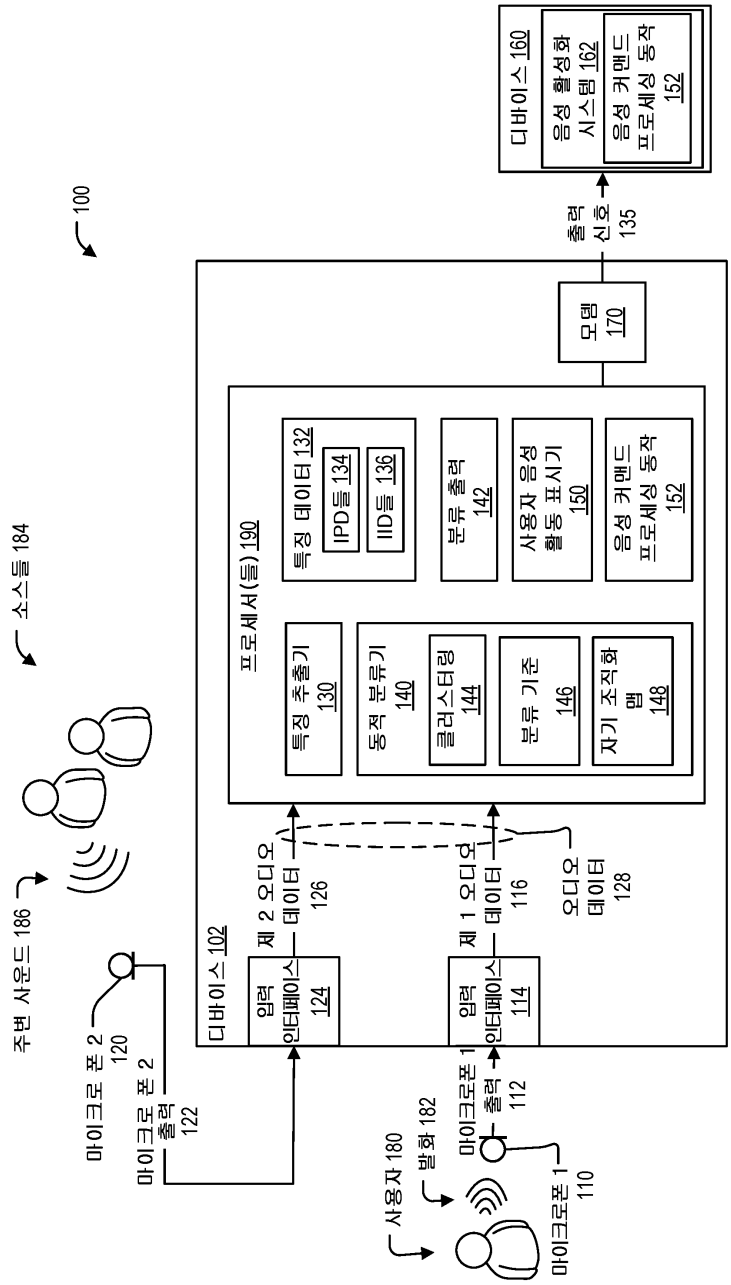
[0196] 당업자는 또한, 본원에 개시된 구현들과 관련하여 설명된 다양한 예시적인 논리 블록들, 구성들, 모듈들, 회로들, 및 알고리즘 단계들이 전자 하드웨어, 프로세서에 의해 실행되는 컴퓨터 소프트웨어, 또는 이 양자의 조합으로 구현될 수도 있음을 인식할 것이다. 다양한 예시적인 컴포넌트들, 블록들, 구성들, 모듈들, 회로들, 및 단계들이 일반적으로 그들의 기능성의 관점에서 위에서 설명되었다. 그러한 기능성이 하드웨어 또는 프로세서 실행가능한 명령들로 구현될지 여부는, 전체 시스템에 부과된 설계 제약 및 특정 애플리케이션에 의존한다. 당업자들은 각각의 특정 애플리케이션에 대해 다양한 방식들로 설명된 기능성을 구현할 수도 있으며, 이러한 구현 판정들은 본 개시의 범위로부터 벗어남을 야기하는 것으로서 해석되어서는 안된다.

[0197] 본 명세서에 개시된 구현들과 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어에서, 프로세서에 의해 실행되는 소프트웨어 모듈에서, 또는 이들 양자의 조합에서 직접 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, 착탈식 디스크, 컴팩트 디스크 판독 전용 메모리 (CD-ROM), 또는 당업계에 알려져 있는 임의의 다른 형태의 비일시적 저장 매체에 상주할 수도 있다. 예시적인 저장 매체는, 프로세서가 저장 매체로부터 정보를 판독하고 저장 매체에 정보를 기입할 수도 있도록 프로세서에 커플링된다. 대안으로, 저장 매체는 프로세서에 통합될 수도 있다. 프로세서 및 저장 매체는 주문형 집적 회로 (ASIC)에 상주할 수도 있다. ASIC 은 컴퓨팅 디바이스 또는 사용자 단말기에 상주할 수도 있다. 대안으로, 프로세서 및 저장 매체는 컴퓨팅 디바이스 또는 사용자 단말기에서 이산 컴포넌트들로서 상주할 수도 있다.

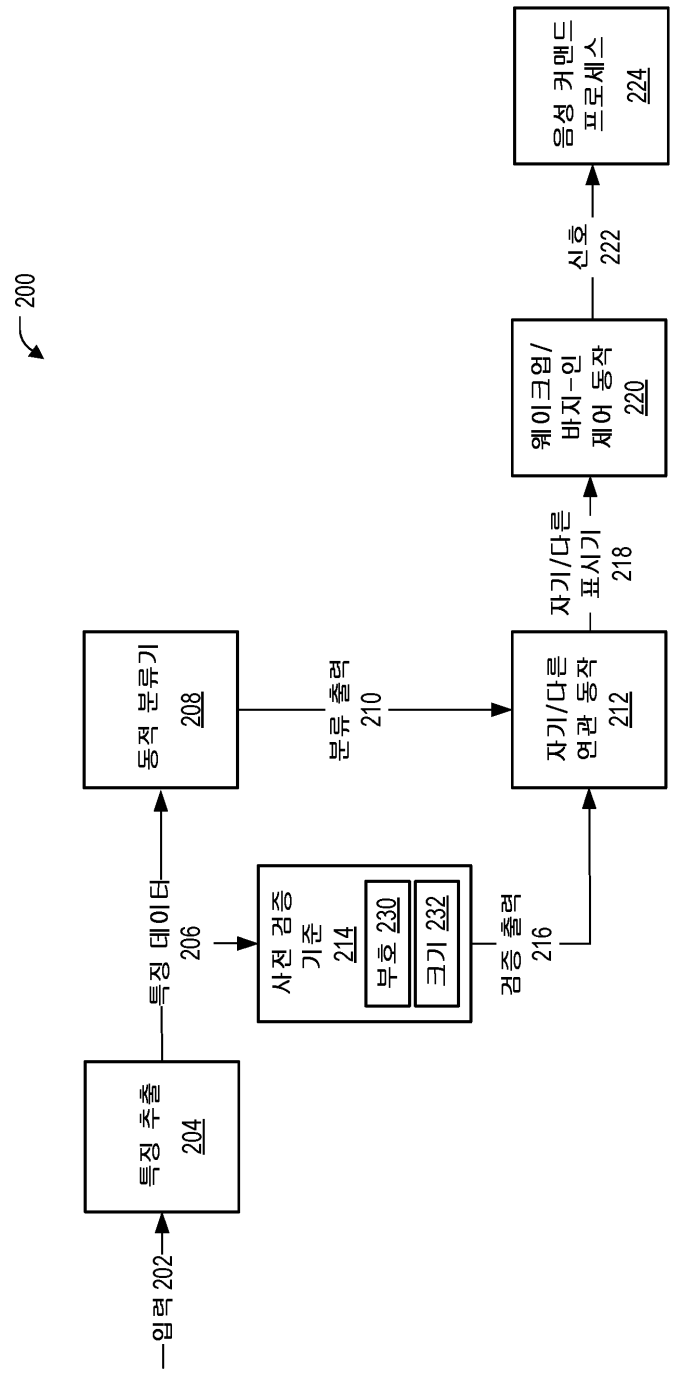
[0198] 개시된 실시형태들의 상기 설명은 당업자로 하여금 개시된 실시형태들을 제조 또는 이용할 수 있도록 제공된다. 이들 양태들에 대한 여러 변경들은 당업자들에게 명백할 것이며, 본원에서 정의된 원리들은 본 개시물의 범위로부터 일탈함이 없이 다른 양태들에 적용될 수도 있다. 따라서, 본 개시물은 본원에서 나타낸 양태들에 한정하려는 것이 아니라, 다음 청구항들에 의해 정의되는 바와 같은 원리들 및 신규한 특징들과 가능한 부합하는 최광의의 범위를 부여하려는 것이다.

도면

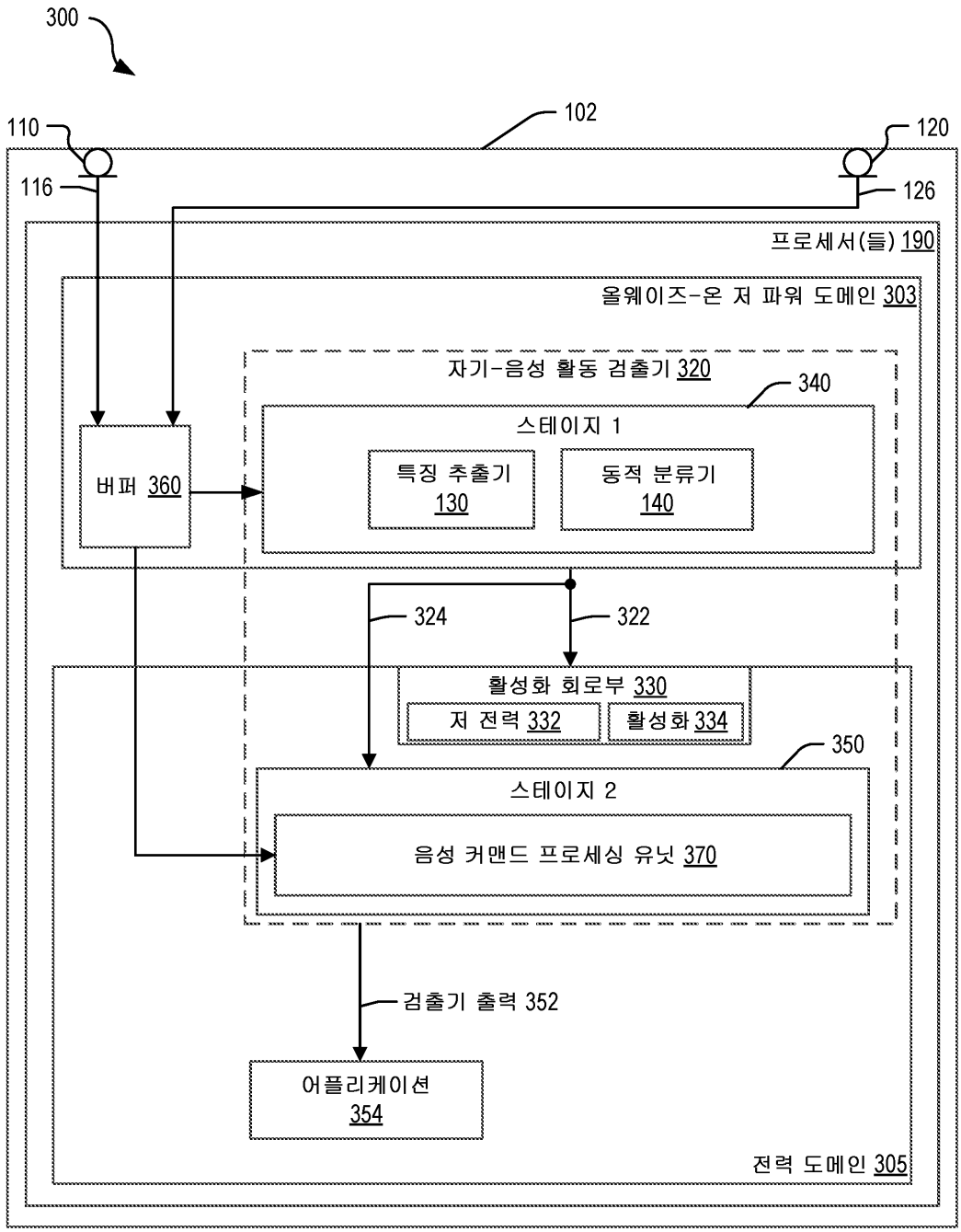
도면1


100
소스들 184
주변 사운드 186
사용자 180
발화 182
마이크로 폰 2
120
마이크로 폰 2
출력
122
마이크로폰 1
110
마이크로폰 1
출력
112
디바이스 102
입력
인터페이스
124
입력
인터페이스
114
제 2 오디오
데이터
126
제 1 오디오
데이터
116
오디오
데이터
128
프로세서(들) 190
특징 추출기
130
동적 분류기
140
클러스터링
144
분류 기준
146
자기 조직화
맵
148
특징 데이터 132
IPD들 134
IID들 136
분류 출력
142
사용자 음성
활동 표시기
150
음성 커맨드
프로세싱 동작
152
모뎀
170
출력
신호
135
디바이스 160
음성 활성화
시스템 162
음성 커맨드
프로세싱 동작
152

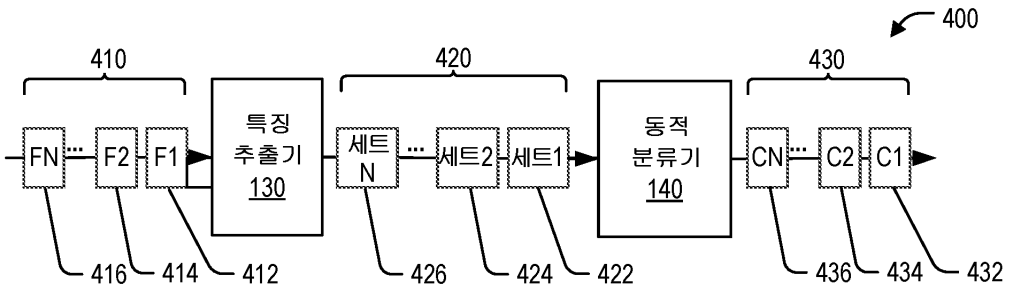
도면2


200
입력 202
특징 추출 204
특징 데이터 206
동적 분류기 208
분류 출력 210
사전 검증 기준 214
부호 230
크기 232
검증 출력 216
자기/다른 연관 동작 212
자기/다른 표시기 218
웨이크업/바지-인 제어 동작 220
신호 222
음성 커맨드 프로세스 224

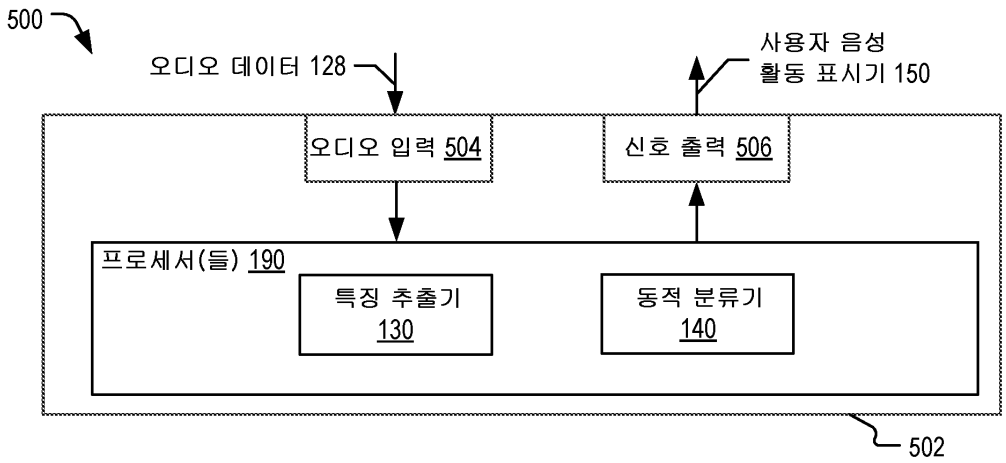
도면3


300
102
110
120
116
126
프로세서(들) 190
올웨이즈-온 저 파워 도메인 303
자기-음성 활동 검출기 320
340
스테이지 1
버퍼 360
특징 추출기 130
동적 분류기 140
324
322
활성화 회로부 330
저 전력 332
활성화 334
350
스테이지 2
음성 커맨드 프로세싱 유닛 370
검출기 출력 352
어플리케이션 354
전력 도메인 305

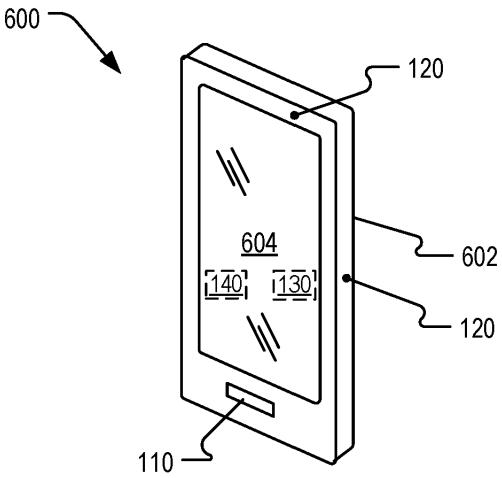
도면4


400
410
420
430
FN
F2
F1
특징 추출기 130
세트 N
세트2
세트1
동적 분류기 140
CN
C2
C1
416
414
412
426
424
422
436
434
432

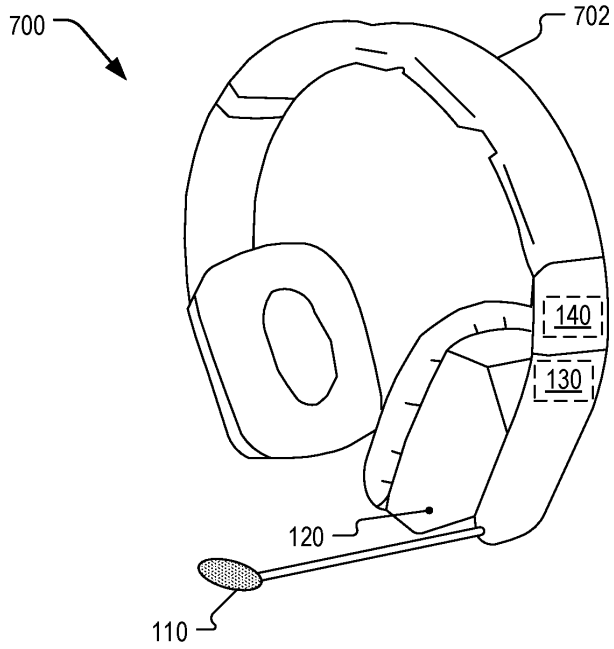
도면5


500
오디오 데이터 128
사용자 음성
활동 표시기 150
오디오 입력 504
신호 출력 506
프로세서(들) 190
특징 추출기
130
동적 분류기
140
502

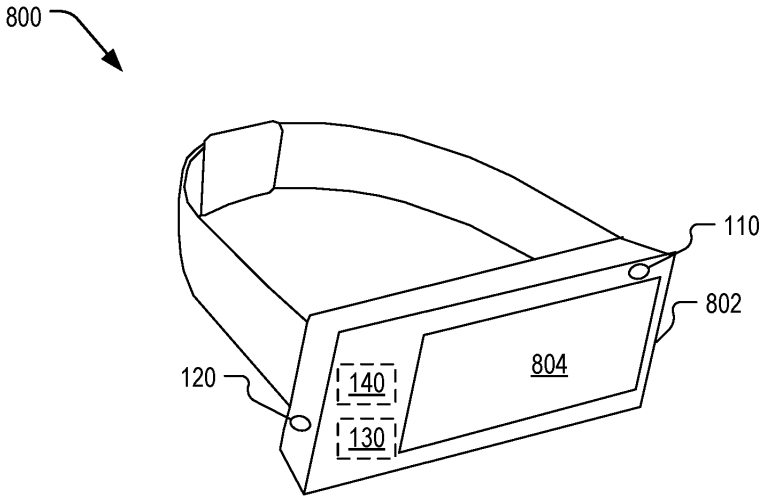
도면6


600
120
604
140
130
602
120
110

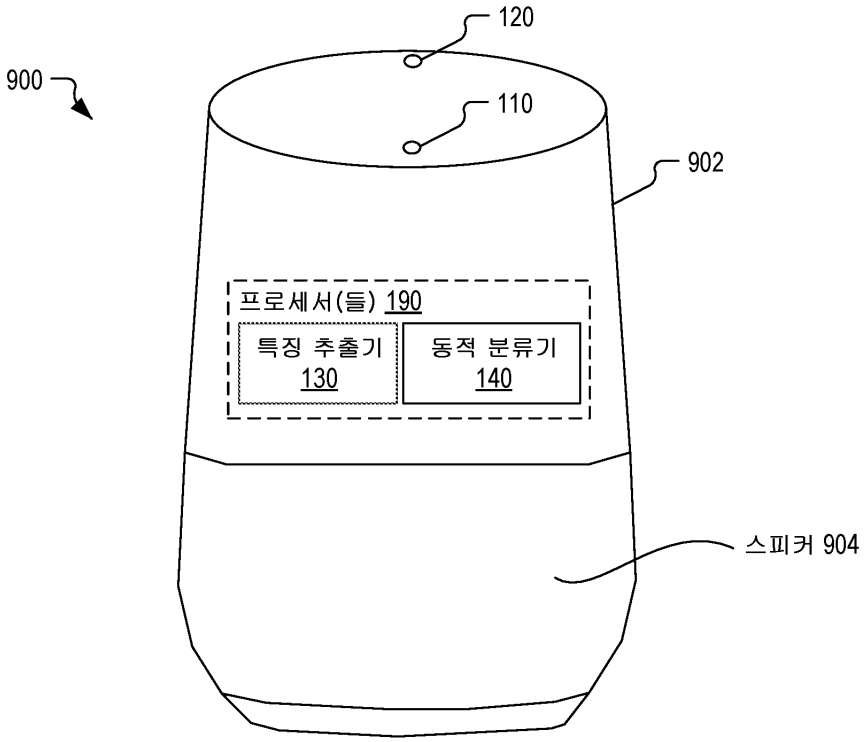
도면7



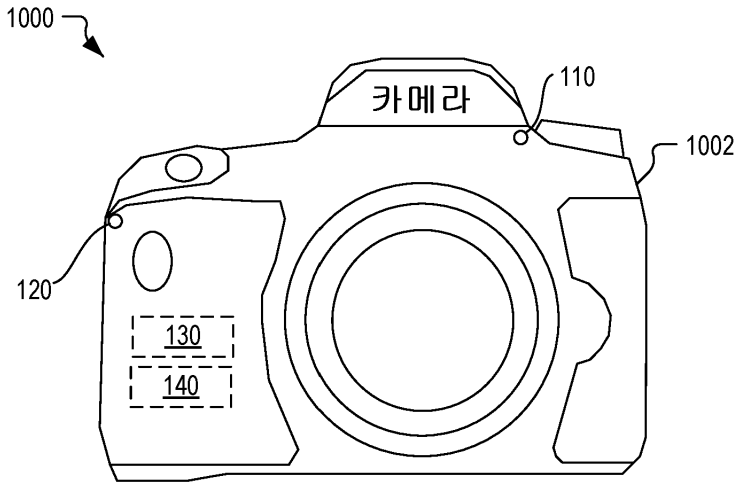
도면8



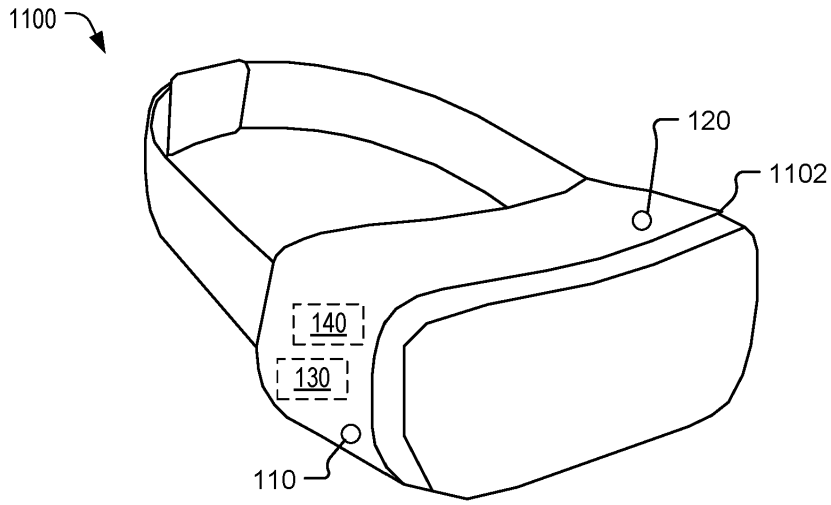
도면9


900
120
110
902
프로세서(들) 190
특징 추출기
130
동적 분류기
140
스피커 904

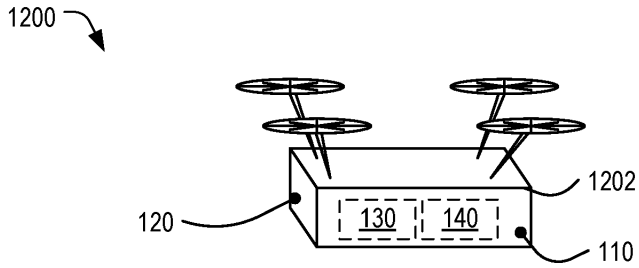
도면10


1000
카메라
110
1002
120
130
140

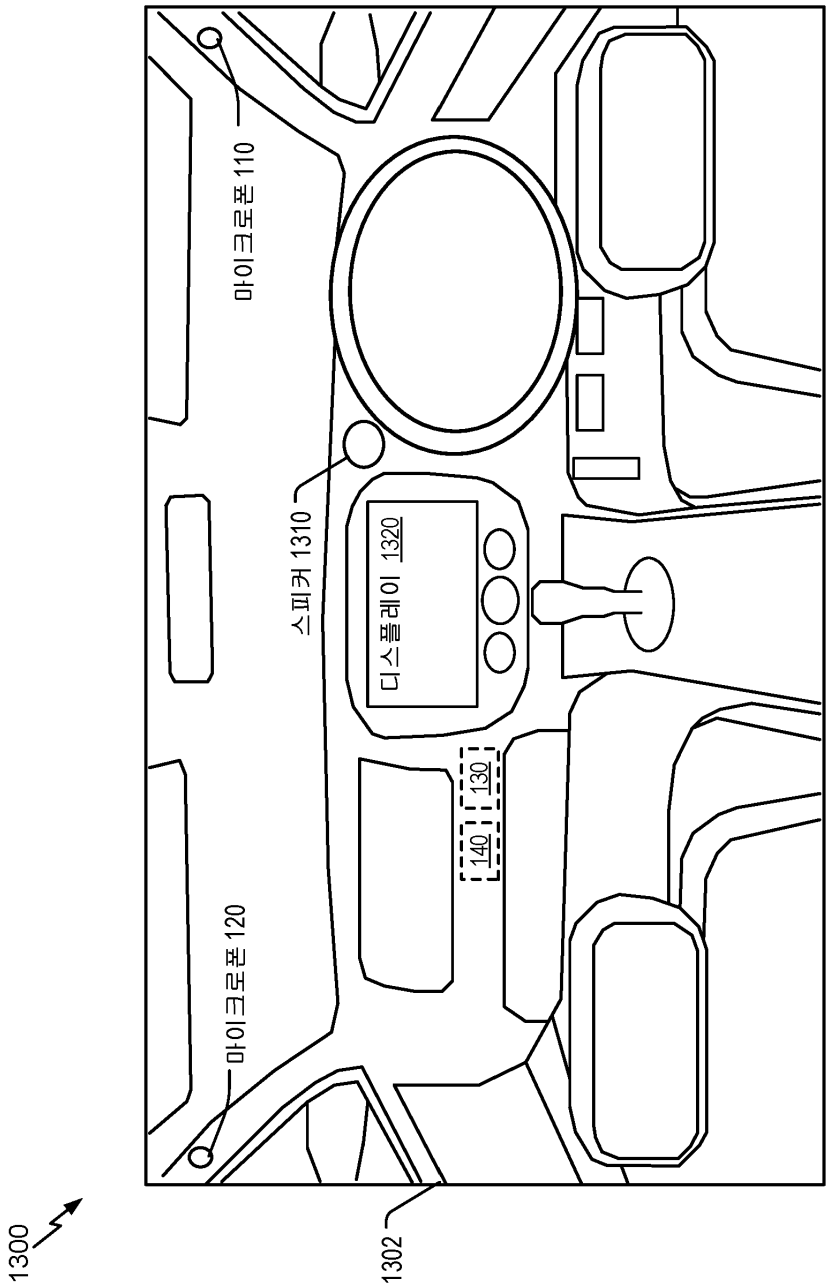
도면11



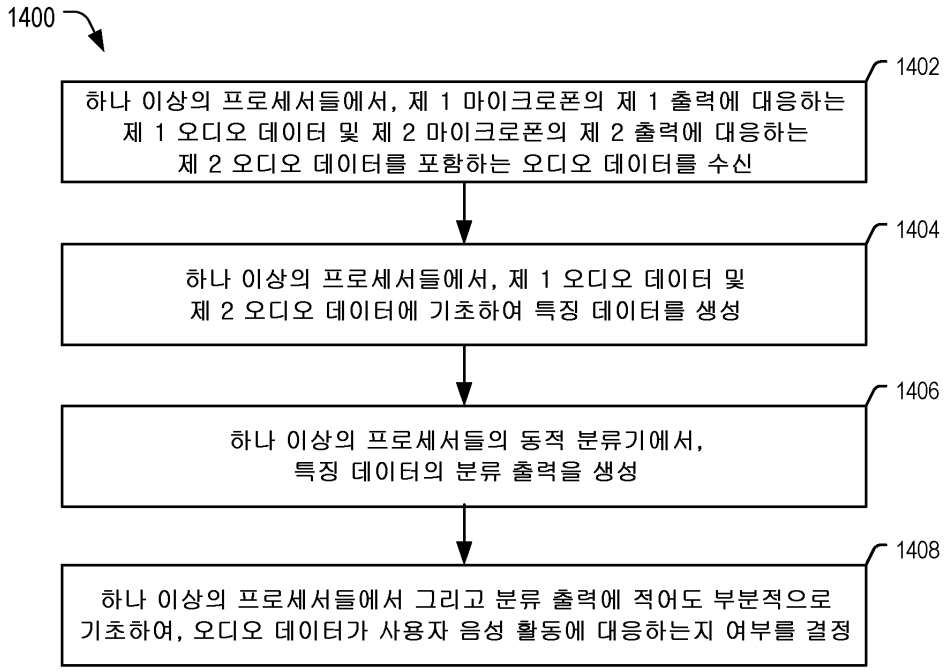
도면12



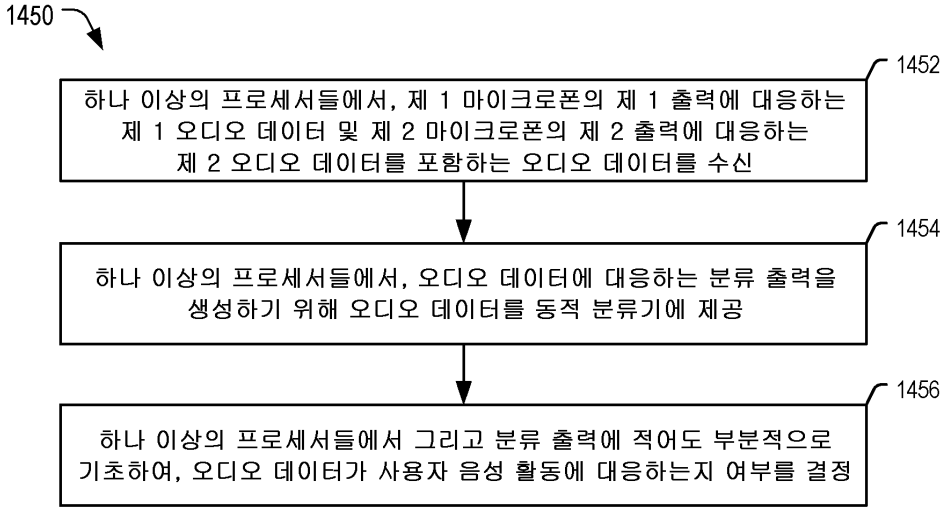
도면13



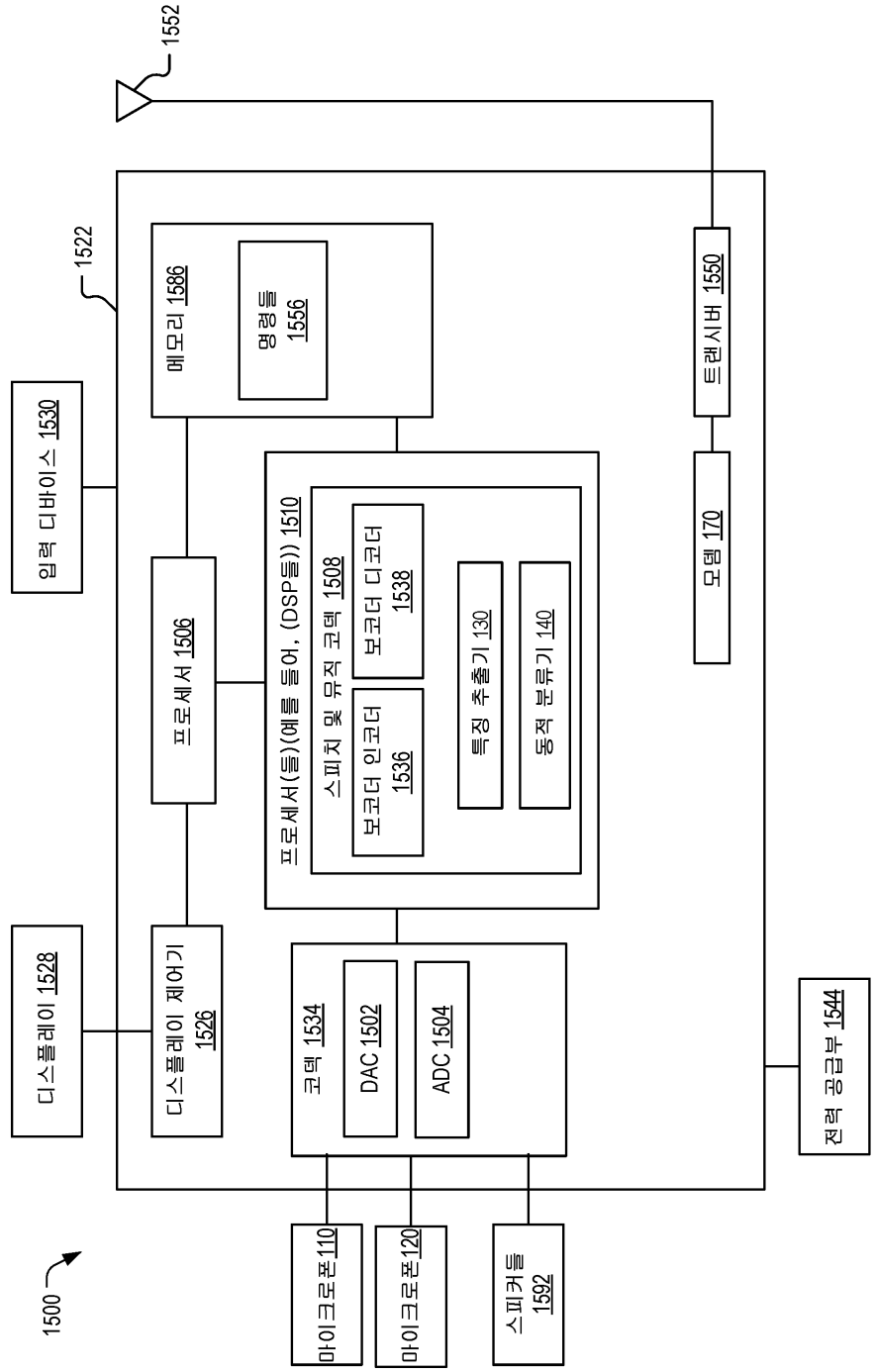
도면14a


1400
1402
하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신
1404
하나 이상의 프로세서들에서, 제 1 오디오 데이터 및 제 2 오디오 데이터에 기초하여 특징 데이터를 생성
1406
하나 이상의 프로세서들의 동적 분류기에서, 특징 데이터의 분류 출력을 생성
1408
하나 이상의 프로세서들에서 그리고 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정

도면14b


1450
1452
하나 이상의 프로세서들에서, 제 1 마이크로폰의 제 1 출력에 대응하는 제 1 오디오 데이터 및 제 2 마이크로폰의 제 2 출력에 대응하는 제 2 오디오 데이터를 포함하는 오디오 데이터를 수신
1454
하나 이상의 프로세서들에서, 오디오 데이터에 대응하는 분류 출력을 생성하기 위해 오디오 데이터를 동적 분류기에 제공
1456
하나 이상의 프로세서들에서 그리고 분류 출력에 적어도 부분적으로 기초하여, 오디오 데이터가 사용자 음성 활동에 대응하는지 여부를 결정

도면15


1552
1522
메모리 1586
명령들 1556
입력 디바이스 1530
프로세서 1506
프로세서(들)(예를 들어, (DSP들)) 1510
스피치 및 뮤직 코덱 1508
보코더 디코더 1538
보코더 인코더 1536
특징 추출기 130
동적 분류기 140
트랜시버 1550
모뎀 170
디스플레이 1528
디스플레이 제어기 1526
코덱 1534
DAC 1502
ADC 1504
전력 공급부 1544
1500
마이크로폰 110
마이크로폰 120
스피커들 1592