



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0062392
(43) 공개일자 2020년06월03일

(51) 국제특허분류(Int. Cl.)
G10L 17/24 (2013.01) G06F 3/16 (2018.01)
G10L 21/16 (2013.01)
(52) CPC특허분류
G10L 17/24 (2013.01)
G06F 3/167 (2013.01)
(21) 출원번호 10-2020-7015307(분할)
(22) 출원일자(국제) 2014년02월07일
심사청구일자 2020년05월27일
(62) 원출원 특허 10-2018-7017535
원출원일자(국제) 2014년02월07일
심사청구일자 2018년06월27일
(85) 번역문제출일자 2020년05월27일
(86) 국제출원번호 PCT/US2014/015418
(87) 국제공개번호 WO 2014/124332
국제공개일자 2014년08월14일
(30) 우선권주장
61/762,260 2013년02월07일 미국(US)

(71) 출원인
애플 인크.
미국 캘리포니아 (우편번호 95014) 쿠퍼티노 원
애플 파크 웨이
(72) 발명자
바인더, 저스틴, 지.
미국 95014 캘리포니아주 쿠퍼티노 인피니트 루프
1 애플인크. 내
택킨, 오누르
미국 95014 캘리포니아주 쿠퍼티노 인피니트 루프
1 애플인크. 내
(뒷면에 계속)
(74) 대리인
장덕순, 백만기

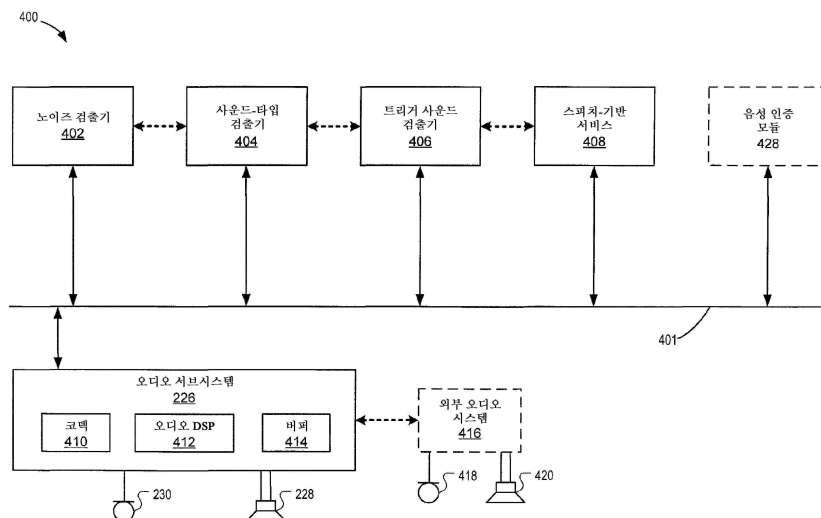
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 디지털 어시스턴트를 위한 음성 트리거

(57) 요약

본 명세서에는 음성 트리거를 동작시키기 위한 방법이 제공되었다. 일부 구현예들에서, 이 방법은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다. 이 방법은 사운드 입력을 수신하는 단계를 포함한다. 사운드 입력은 구술된 단어 또는 구, 또는 이들의 일부분에 대응할 수 있다. 이 방법은 사운드 입력의 적어도 일부분이 사람의 음성과 같은 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 단계를 포함한다. 이 방법은, 사운드 입력의 적어도 일부분이 사전결정된 타입에 대응한다는 결정에 따라, 사운드 입력이 사전결정된 트리거 단어 또는 구와 같은 사전결정된 콘텐츠를 포함하는지 여부를 결정하는 단계를 포함한다. 이 방법은 또한, 사운드 입력이 사전결정된 콘텐츠를 포함한다는 결정에 따라, 음성-기반 디지털 어시스턴트와 같은 스피치-기반 서비스를 개시하는 단계를 포함한다.

대표도 - 도4



(52) CPC특허분류

G10L 21/16 (2013.01)

(72) 발명자

포스트, 사무엘, 디.

미국 95014 캘리포니아주 쿠퍼티노 인피니트 루프
1 애플링크. 내

그루버, 토마스, 알.

미국 95014 캘리포니아주 쿠퍼티노 인피니트 루프
1 애플링크. 내

명세서

청구범위

청구항 1

하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행되는, 음성 트리거(voice trigger)를 동작시키는 방법으로서,

사운드 입력을 수신하는 단계;

상기 사운드 입력의 입력 표현을 상기 음성 트리거를 위한 하나 이상의 트리거 단어의 하나 이상의 기준 표현에 비교하는 것에 기초하여 상기 사운드 입력이 사전결정된 콘텐츠에 대응하는지 여부를 결정하는 단계;

상기 사운드 입력이 사전결정된 콘텐츠에 대응하지 않는다고 결정하는 것에 따라, 스피치 기반 서비스를 개시하는 것을 중지하는 단계;

상기 사운드 입력이 사전결정된 콘텐츠에 대응하지 않는다고 결정한 이후에, 상기 사운드 입력을 수신하는 사전결정된 지속시간 내에, 상기 스피치 기반 서비스를 개시하기 위해 사용자 입력을 검출하는 단계; 및

상기 스피치 기반 서비스를 개시하기 위해 상기 사용자 입력을 검출하는 것에 따라, 상기 하나 이상의 기준 표현과 상기 사운드 입력의 상기 입력 표현 사이의 부합(match)의 신뢰도 레벨(confidence level)을 증가시키기 위해 상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계

를 포함하는 방법.

청구항 2

제1항에 있어서,

상기 하나 이상의 기준 표현을 조정하는 단계는 상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현의 이동 평균을 업데이트하는 단계를 포함하는, 방법.

청구항 3

제1항에 있어서,

상기 스피치 기반 서비스를 개시하기 위한 상기 사용자 입력은, 상기 전자 디바이스의 버튼의 사용자 선택 또는 상기 전자 디바이스에 의해 디스플레이되는 어포던스(affordance)의 사용자 선택에 대응하는, 방법.

청구항 4

제1항에 있어서,

상기 사운드 입력을 수신하기 이전에, 복수의 사운드 입력을 수신하는 단계 - 상기 복수의 사운드 입력 중 각각의 사운드 입력은 상기 하나 이상의 트리거 단어를 포함함 -; 및

상기 복수의 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 생성하는 단계

를 더 포함하는 방법.

청구항 5

제1항에 있어서,

상기 사운드 입력이 상기 사전결정된 콘텐츠에 대응한다고 결정하는 것에 따라, 상기 스피치 기반 서비스를 개시하는 단계

를 더 포함하는 방법.

청구항 6

제5항에 있어서,

상기 사운드 입력의 상기 입력 표현을 상기 하나 이상의 기준 표현에 비교하는 것은 상기 사운드 입력의 상기 입력 표현이 상기 하나 이상의 기준 표현에 사전결정된 신뢰도 레벨로 부합하는지 여부를 결정하는 것을 포함하고,

상기 방법은, 상기 스피치 기반 서비스를 개시하는 것에 따라:

상기 입력 표현이 상기 하나 이상의 기준 표현에 상기 사전결정된 신뢰도 레벨로 부합한다는 결정에 따라:

상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계; 및

상기 사운드 입력의 상기 입력 표현이 상기 하나 이상의 기준 표현에 상기 사전결정된 신뢰도 레벨로 부합하지 않는다는 결정에 따라:

상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 것을 중지하는 단계

를 더 포함하는, 방법.

청구항 7

제1항에 있어서,

상기 사운드 입력이 상기 사전결정된 콘텐츠에 대응하지 않는다고 결정한 이후에, 상기 사운드 입력에 연속하는 하나 이상의 사운드 입력을 수신하는 단계 - 상기 하나 이상의 사운드 입력 중 각각의 사운드 입력은 상기 하나 이상의 트리거 단어를 포함함 -; 및

상기 하나 이상의 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계

를 더 포함하는 방법.

청구항 8

제7항에 있어서,

상기 하나 이상의 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계는 상기 하나 이상의 사운드 입력이 적어도 2개의 연속적인 사운드 입력을 포함한다는 결정에 따라 수행되는, 방법.

청구항 9

제7항에 있어서,

상기 하나 이상의 사운드 입력은 적어도 2개의 사운드 입력을 포함하고, 상기 하나 이상의 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계는 상기 적어도 2개의 사운드 입력이 사전결정된 기간 내에 수신되었다는 결정에 따라 수행되는, 방법.

청구항 10

제1항에 있어서,

상기 사운드 입력의 상기 입력 표현을 상기 하나 이상의 기준 표현에 비교하는 것은 알고리즘을 이용하여 상기 입력 표현을 상기 하나 이상의 기준 표현에 비교하는 것을 포함하고,

상기 방법은,

상기 스피치 기반 서비스를 개시하기 위해 상기 사용자 입력을 검출하는 것에 따라, 상기 알고리즘을 조정하는 단계

를 더 포함하는, 방법.

청구항 11

제1항에 있어서,

상기 사운드 입력이 사전결정된 콘텐츠에 대응하는지 여부를 결정하는 단계는 제1 프로세서를 이용하여 수행되

고,

상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계는 제2 프로세서를 이용하여 수행되는, 방법.

청구항 12

제11항에 있어서,

상기 제2 프로세서는 상기 사운드 입력이 사전결정된 콘텐츠에 대응하는지 여부를 결정할 때 스탠바이 모드에 있는, 방법.

청구항 13

제11항에 있어서,

상기 사운드 입력을 상기 전자 디바이스의 메모리에 저장하는 단계; 및

상기 스피치 기반 서비스를 개시하기 위해 상기 사용자 입력을 검출하는 것에 따라, 상기 제2 프로세서를 활성화하는 단계

를 더 포함하고,

상기 사운드 입력을 이용하여 상기 하나 이상의 기준 표현을 조정하는 단계는, 상기 활성화된 제2 프로세서에 의해, 상기 사운드 입력에 액세스하는 단계를 포함하는, 방법.

청구항 14

제1항에 있어서,

상기 사운드 입력이 사전결정된 콘텐츠에 대응하는지 여부를 결정하는 단계 이전에, 상기 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 단계

를 더 포함하는 방법.

청구항 15

제14항에 있어서,

상기 사운드 입력이 사전결정된 콘텐츠에 대응하는지 여부를 결정하는 단계는 상기 사운드 입력이 상기 사전결정된 타입의 사운드에 대응한다고 결정하는 것에 따라 수행되는, 방법.

청구항 16

제1항에 있어서,

상기 스피치 기반 서비스는 디지털 어시스턴트를 포함하는, 방법.

청구항 17

전자 디바이스의 하나 이상의 프로세서에 의해 실행되기 위한 하나 이상의 프로그램을 저장하는 컴퓨터 판독가능한 저장 매체로서, 상기 하나 이상의 프로그램은 제1항 내지 제16항 중 어느 한 항에 기재된 방법을 수행하기 위한 명령어들을 포함하는, 컴퓨터 판독가능한 저장 매체.

청구항 18

전자 디바이스로서,

하나 이상의 프로세서;

메모리; 및

하나 이상의 프로그램을 포함하고,

상기 하나 이상의 프로그램은 상기 메모리에 저장되고 상기 하나 이상의 프로세서에 의해 실행되도록 구성되며,

상기 하나 이상의 프로그램은 제1항 내지 제16항 중 어느 한 항에 기재된 방법을 수행하기 위한 명령어들을 포함하는, 전자 디바이스.

발명의 설명

기술 분야

[0001] 관련 출원 상호 참조

[0002] 본 출원은 2013년 2월 7일 출원된 미국 임시출원 제61/762,260호, 발명의 명칭 "VOICE TRIGGER FOR A DIGITAL ASSISTANT"의 우선권을 주장하며, 이것은 전체가 본 명세서에 참조로서 포함된다.

[0003] 개시된 구현에는 일반적으로 디지털 어시스턴트에 관한 것으로, 보다 구체적으로는 디지털 어시스턴트를 위한 음성 트리거 방법 및 시스템에 관한 것이다.

배경 기술

[0004] 최근에, 애플(Apple)사의 시리(SIRI)와 같은 음성-기반 디지털 어시스턴트가 웹 탐색 및 내비게이션과 같은 다양한 작업들을 다루기 위한 시장에 도입되었다. 이러한 음성-기반 디지털 어시스턴트의 일 장점은 사용자들이 디바이스를 다루거나 볼 필요조차도 없이 핸드-프리 방식으로 디바이스와 상호작용할 수 있다는 것이다. 핸드-프리 동작은, 사용자들이 운전을 할 때와 같이 사용자가 디바이스를 물리적으로 다룰 수 없거나 다루어서는 안 되는 경우에서 특히 바람직할 수 있다. 그러나, 음성-기반 어시스턴트를 개시하기 위해서, 사용자들은 통상적으로 버튼을 누르거나 터치 스크린 상의 아이콘을 선택해야만 한다. 이러한 접촉식 입력은 핸드-프리 경험을 손상시킨다. 따라서, 접촉식 입력이 아닌 음성 입력 또는 신호를 이용하여 음성-기반 디지털 어시스턴트(또는 다른 스피치-기반 서비스)를 활성화하는 방법 및 시스템을 제공하는 것이 바람직할 수 있다.

[0005] 음성 입력을 이용하여 음성-기반 어시스턴트를 활성화하는 것은 음성 입력을 검출하기 위해 오디오 채널을 모니터링하는 것을 필요로 한다. 이러한 모니터링은 배터리에 의존하고 이러한 음성-기반 디지털 어시스턴트들이 주로 실행되는 핸드헬드 또는 휴대용 디바이스 상의 제한된 리소스인 전력을 소비한다. 따라서, 디바이스 상에서 음성- 및/또는 스피치-기반 서비스들을 개시하도록 사용될 수 있는 에너지 효율적인 음성 트리거를 제공하는 것이 바람직할 것이다.

발명의 내용

[0006] 따라서, 제한된 전력 리소스를 과도하게 소비하지 않고 "항상-청취하는" 음성 트리거 기능을 제공할 수 있는 저 전력 음성 트리거에 대한 필요성이 존재한다. 아래에서 기술되는 구현예들은 전자 디바이스에서 음성 트리거를 이용하여 음성-기반 어시스턴트를 개시하기 위한 시스템들 및 방법들을 제공한다. 음성-기반 디지털 어시스턴트(또는 스피치-텍스트(speech-to-text) 전사 서비스와 같은 다른 스피치-기반 서비스)와의 상호작용들은 종종 사용자가 디지털 어시스턴트를 활성화하기 위해 디바이스 상의 어포던스(affordance)(예를 들어, 버튼 또는 아이콘)를 누를 때 시작되고, 그에 이어 디바이스가 광, 사운드(예를 들어, 삐 소리), 또는 음성화된 출력(예를 들어, "무엇을 해드릴까요?")과 같이 디지털 어시스턴트가 활동중이며 청취중이라는 일부 표시를 사용자에게 제공한다. 본 명세서에서 기술되는 바와 같이, 음성 트리거는 또한 사용자에게 의한 물리적 상호작용을 요구하지 않고 특정한 사전결정된 단어, 구(phrase), 또는 사운드에 응답하여 활성화되도록 구현될 수도 있다. 예를 들어, 사용자는 구 "안녕, 시리"라고 말함으로써 아이폰(IPHONE) 상의 시리 디지털 어시스턴트(둘 모두가 본 출원의 양수인인 애플 인크.(Apple Inc.)에 의해 제공됨)를 활성화할 수 있다. 그에 응답하여, 디바이스는 사용자에게 청취 모드가 활동중임을 나타내는 삐 소리, 사운드, 또는 스피치 출력(예를 들어, "무엇을 해드릴까요?")을 출력한다. 따라서, 사용자는 디지털 어시스턴트 기능을 제공하는 디바이스를 물리적으로 터치해야 할 필요 없이 디지털 어시스턴트와의 상호작용을 개시할 수 있다.

[0007] 음성 트리거를 이용하여 스피치-기반 서비스를 개시하기 위한 일 기술은 사전결정된 트리거 단어, 구, 또는 사운드(이들 중 임의의 것이 본 명세서에서 "트리거 사운드"로 지칭될 수 있다)를 계속해서 청취하는 스피치-기반 서비스를 갖는 것이다. 그러나, 스피치-기반 서비스(예를 들어, 음성-기반 디지털 어시스턴트)를 계속해서 동작시키는 것은 상당한 오디오 프로세싱 및 배터리 전력을 요구한다. 음성 트리거 기능을 제공함으로써 소비되는 전력을 감소시키기 위해서, 몇몇 기술들이 사용될 수 있다. 일부 구현예들에서, 전자 디바이스의 메인 프로세서(즉, "애플리케이션 프로세서")가 저전력 또는 전력공급되지 않는 상태로 유지되는 반면, (예를 들어, 애플리케이션 프로세서에 의존하지 않기 때문에) 더 적은 전력을 사용하는 하나 이상의 사운드 검출기는 활동적으로

남아있게 된다. (저전력 또는 전력공급되지 않는 상태에 있는 경우, 애플리케이션 프로세서 또는 임의의 다른 프로세서, 프로그램, 또는 모듈이 비활동적이거나 스탠바이 모드에 있는 것으로 기술될 수 있다.) 예를 들어, 애플리케이션 프로세서가 비활동적일 때에도 저전력 사운드 검출기가 트리거 사운드에 대해 오디오 채널을 모니터링하도록 사용된다. 이러한 사운드 검출기는 때때로 본 명세서에서 트리거 사운드 검출기로서 지칭된다. 일부 구현예들에서, 이것은 특정한 사운드, 음소(phoneme), 및/또는 단어를 검출하도록 구성된다. (하드웨어 및/또는 소프트웨어 컴포넌트들을 포함하는) 트리거 사운드 검출기는 특정한 단어, 사운드, 또는 구를 인식하도록 설계되지만, 일반적으로 이러한 작업이 더 큰 컴퓨터 및 전력 리소스를 요구하기 때문에, 전체 스피치-텍스트 기능을 제공하기 위해 최적화될 수 없다. 따라서, 일부 구현예들에서, 트리거 사운드 검출기는 음성 입력이 사전정의된 패턴(예를 들어, 단어들 "안녕, 시리"에 부합하는 소리 패턴)을 포함하는지 여부를 인식하지만, 음성 입력을 텍스트로 변환하거나 상당한 양의 다른 단어들을 인식할 수 없다(또는 그렇게 구성되지 않는다). 트리거 사운드가 검출되면, 그 다음 디지털 어시스턴트는 사용자가 음성 커맨드를 제공할 수 있도록 스탠바이 모드 밖으로 나가게 된다.

[0008] 일부 구현예들에서, 트리거 사운드 검출기는 단어들의 세트, 구, 사운드, 및/또는 이들의 조합과 같은 몇몇 서로 다른 트리거 사운드들을 검출하도록 구성된다. 그 다음 사용자는 스피치-기반 서비스를 개시하기 위해서 임의의 이러한 사운드들을 사용할 수 있다. 일 예에서, 음성 트리거는 "안녕, 시리", "일어나, 시리", "나의 디지털 어시스턴트를 호출해줘", 또는 "안녕, 헬(HAL), 읽고 있니 헬?"과 같은 구들에 대해 응답하도록 사전구성된다. 일부 구현예들에서, 사용자는 사전구성된 트리거 사운드들 중 하나를 단독 트리거 사운드로서 선택해야만 한다. 일부 구현예들에서, 사용자는 사전구성된 트리거 사운드들의 하위세트를 선택하며, 그에 따라 사용자가 서로 다른 트리거 사운드들을 이용하여 스피치-기반 서비스를 개시할 수 있다. 일부 구현예들에서, 사전구성된 트리거 사운드들 전부가 유효 트리거 사운드들로 남아있다.

[0009] 일부 구현예들에서, 트리거 사운드 검출기가 대부분의 시간 동안 저전력 또는 전력공급되지 않는 모드로 유지될 수 있도록 다른 사운드 검출기가 사용된다. 예를 들어, 상이한 타입의 사운드 검출기(예를 들어, 트리거 사운드 검출기보다 더 적은 전력을 사용하는 사운드 검출기)가 사운드 입력이 소정 타입의 사운드에 대응하는지 여부를 결정하기 위해 오디오 채널을 모니터링하도록 사용된다. 사운드들은 소정의 식별가능한 사운드들의 특징에 기초하여 상이한 "타입들"로 분류된다. 예를 들어, "사람의 음성" 타입의 사운드는 소정의 스펙트럼 콘텐츠, 주기성, 본질적 주파수 등을 갖는다. 다른 타입의 사운드들(예를 들어, 휘파람, 손뼉 등)은 다른 특징들을 갖는다. 서로 다른 타입의 사운드들은 본 명세서에 기술된 것과 같은 오디오 및/또는 신호 프로세싱 기술들을 이용하여 식별된다. 이러한 사운드 검출기는 때때로 본 명세서에서 "사운드-타입 검출기"로 지칭된다. 예를 들어, 만약 사전결정된 트리거 구가 "안녕, 시리"라면, 사운드-타입 검출기는 입력이 사람의 스피치에 대응할 가능성이 높은지 여부를 결정한다. 만약 트리거 사운드가 휘파람과 같은 비-음성 사운드라면, 사운드-타입 검출기는 사운드 입력이 휘파람에 대응할 가능성이 높은지 여부를 결정한다. 적절한 타입의 사운드가 검출되면, 사운드-타입 검출기는 사운드를 추가로 프로세싱 및/또는 분석하기 위해 트리거 사운드 검출기를 개시한다. 그리고 (예를 들어, 사운드-타입 검출기가 트리거 사운드 검출기보다 더 낮은 전력 수요 및/또는 더욱 효율적인 오디오 프로세싱 알고리즘을 갖는 회로를 사용하기 때문에) 사운드-타입 검출기가 트리거 사운드 검출기보다 더 적은 전력을 요구하므로, 음성 트리거 기능은 트리거 사운드 검출기를 단독으로 사용하는 것보다도 더 적은 전력을 소비한다.

[0010] 일부 구현예들에서, 전술된 사운드-타입 검출기 및 트리거 사운드 검출기 모두가 대부분의 시간 동안 저전력 또는 전력공급되지 않는 모드로 유지될 수 있도록 또 다른 사운드 검출기가 사용된다. 예를 들어, 사운드-타입 검출기보다 더 적은 전력을 사용하는 사운드 검출기는 사운드 입력이 진폭(예를 들어, 볼륨) 임계값과 같은 사전결정된 조건을 만족시키는지 여부를 결정하기 위해 오디오 채널을 모니터링하도록 사용된다. 이러한 사운드 검출기는 본 명세서에서 노이즈 검출기로 지칭될 수 있다. 노이즈 검출기가 사전결정된 임계값을 만족시키는 사운드를 검출하면, 노이즈 검출기는 사운드를 추가로 프로세싱 및/또는 분석하기 위해 사운드-타입 검출기를 개시한다. 그리고 (예를 들어, 노이즈 검출기가 더 낮은 전력 수요 및/또는 더욱 효율적인 오디오 프로세싱 알고리즘을 갖는 회로를 사용하기 때문에) 노이즈 검출기가 사운드-타입 검출기 또는 트리거 사운드 검출기보다 더 적은 전력을 요구하므로, 음성 트리거 기능은 노이즈 검출기 없이 사운드-타입 검출기와 트리거 사운드 검출기의 조합보다도 더 적은 전력을 소비한다.

[0011] 일부 구현예들에서, 전술된 임의의 하나 이상의 사운드 검출기는, 이들이 "온(on)" 및 "오프(off)" 상태 사이에서 순환되는 듀티 사이클에 따라 동작된다. 이것은 음성 트리거의 전력 소비를 감소시키는 것을 추가로 돕는다. 예를 들어, 일부 구현예들에서, 노이즈 검출기가 10ms 동안 "온"되고(즉, 오디오 채널을 활동적으로

모니터링함) 이어지는 90ms 동안 "오프"된다. 이러한 방식으로, 노이즈 검출기는 시간의 90% 동안 "오프"되는 동시에, 여전히 효율적으로 연속적인 노이즈 검출 기능을 제공한다. 일부 구현예들에서, 사운드 검출기들에 대한 온 및 오프 지속시간은 트리거 사운드가 여전히 입력 중인 동안에 모든 검출기들이 활성화되도록 선택된다. 예를 들어, "안녕, 시리"의 트리거 구에 있어서, 사운드 검출기들은 트리거 구가 듀티 사이클(들) 내의 어디에서 시작하더라도 트리거 사운드 검출기가 충분한 양의 입력을 분석하기 위해 시간 내에 활성화되도록 구성될 수 있다. 예를 들어, 트리거 사운드 검출기는 사운드가 트리거 구에 부합한다고 결정하기에 충분한 사운드인 "녕 시리"를 수신, 프로세싱 및 분석하기 위해 시간 내에 활성화될 것이다. 일부 구현예들에서, 사운드 입력의 더 큰 부분이 분석될 수 있도록 사운드 입력이 업스트림 검출기에 수신되고 전달될 때 메모리 내에 저장된다. 따라서, 트리거 구가 발인된 후까지도 트리거 사운드 검출기가 개시되지 않는다고 해도, 여전히 전체 기록된 트리거 구를 분석할 수 있다.

- [0012] 일부 구현예들은 음성 트리거를 동작하기 위한 방법을 제공한다. 이 방법은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다. 이 방법은 사운드 입력을 수신하는 것을 포함한다. 이 방법은 또한 사운드 입력의 적어도 일부분이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 것을 포함한다. 이 방법은 사운드 입력의 적어도 일부분이 사전결정된 타입에 대응한다는 결정에 따라, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하는 것을 더 포함한다. 이 방법은, 사운드 입력이 사전결정된 콘텐츠를 포함한다는 결정에 따라, 스피치-기반 서비스를 개시하는 것을 더 포함한다. 일부 구현예들에서, 스피치-기반 서비스는 음성-기반 디지털 어시스턴트이다. 일부 구현예들에서, 스피치-기반 서비스는 딕테이션(dictation) 서비스이다.
- [0013] 일부 구현예들에서, 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 것은 제1 사운드 검출기에 의해 수행되고, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하는 것은 제2 사운드 검출기에 의해 수행된다. 일부 구현예들에서, 제1 사운드 검출기는 동작 중에 제2 사운드 검출기보다 더 적은 전력을 소비한다. 일부 구현예들에서, 제1 사운드 검출기는 사운드 입력의 주파수-도메인 분석을 수행한다. 일부 구현예들에서, 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 것은 (예를 들어, 아래에서 논의되는 제3 사운드 검출기에 의해 결정되는 바와 같이) 사운드 입력이 사전결정된 조건을 만족시킨다는 결정에 따라 수행된다.
- [0014] 일부 구현예들에서, 제1 사운드 검출기는 듀티 사이클에 따라 오디오 채널을 주기적으로 모니터링한다. 일부 구현예들에서, 듀티 사이클은 약 20ms의 온-시간 및 약 100ms의 오프-시간을 포함한다.
- [0015] 일부 구현예들에서, 사전결정된 타입은 사람의 음성이고 사전결정된 콘텐츠는 하나 이상의 단어이다. 일부 구현예들에서, 사운드 입력의 적어도 일부분이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 것은 사운드 입력의 적어도 일부분이 사람의 음성의 주파수 특징을 포함하는지 여부를 결정하는 것을 포함한다.
- [0016] 일부 구현예들에서, 제2 사운드 검출기는 사운드 입력이 사전결정된 타입에 대응한다는 제1 사운드 검출기에 의한 결정에 응답하여 개시된다. 일부 구현예들에서, 제2 사운드 검출기는 사운드 입력이 사전결정된 타입에 대응한다는 제1 사운드 검출기에 의한 결정 후 적어도 사전결정된 길이의 시간 동안 동작된다. 일부 구현예들에서, 사전결정된 길이의 시간은 사전결정된 콘텐츠의 지속시간에 대응한다.
- [0017] 일부 구현예들에서, 사전결정된 콘텐츠는 하나 이상의 사전결정된 음소이다. 일부 구현예들에서, 하나 이상의 사전결정된 음소는 적어도 하나의 단어를 구성한다.
- [0018] 일부 구현예들에서, 이 방법은, 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하기 전에, 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하는 것을 포함한다. 일부 구현예들에서, 사전결정된 조건은 진폭 임계값이다. 일부 구현예들에서, 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하는 것은 제3 사운드 검출기에 의해 수행되며, 이때 제3 사운드 검출기는 동작 중에 제1 사운드 검출기보다 더 적은 전력을 소비한다. 일부 구현예들에서, 제3 사운드 검출기는 듀티 사이클에 따라 오디오 채널을 주기적으로 모니터링한다. 일부 구현예들에서, 듀티 사이클은 약 20ms의 온-시간 및 약 500ms의 오프-시간을 포함한다. 일부 구현예들에서, 제3 사운드 검출기는 사운드 입력의 시간-도메인 분석을 수행한다.
- [0019] 일부 구현예들에서, 이 방법은 사운드 입력의 적어도 일부분을 메모리에 저장하고, 스피치-기반 서비스가 개시되면 사운드 입력의 해당 부분을 스피치-기반 서비스에 제공하는 것을 포함한다. 일부 구현예들에서, 사운드 입력의 해당 부분은 직접 메모리 액세스를 이용하여 메모리에 저장된다.
- [0020] 일부 구현예들에서, 이 방법은 사운드 입력이 특정 사용자의 음성에 대응하는지 여부를 결정하는 것을

포함한다. 일부 구현예들에서, 스피치-기반 서비스는 사운드 입력이 사전결정된 콘텐츠를 포함하고 사운드 입력이 특정 사용자의 음성에 대응한다는 결정에 따라 개시된다. 일부 구현예들에서, 스피치-기반 서비스는 사운드 입력이 사전결정된 콘텐츠를 포함하고 사운드 입력이 특정 사용자의 음성에 대응하지 않는다는 결정에 따라 제한된 액세스 모드로 개시된다. 일부 구현예들에서, 이 방법은 사운드 입력이 특정 사용자의 음성에 대응한다는 결정에 따라, 특정 사용자의 이름을 포함하는 음성 프롬프트(prompt)를 출력하는 것을 포함한다.

[0021] 일부 구현예들에서, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하는 것은 사운드 입력의 표현을 기준 표현에 비교하는 것과, 사운드 입력의 표현이 기준 표현에 부합하는 경우 사운드 입력이 사전결정된 콘텐츠를 포함한다고 결정하는 것을 포함한다. 일부 구현예들에서, 만약 사운드 입력의 표현이 사전결정된 신뢰도로 기준 표현에 부합한다면 부합이 결정된다. 일부 구현예들에서, 이 방법은 이 사운드 입력을 포함하는 복수의 사운드 입력을 수신하는 것; 그리고 개별 사운드 입력들이 사전결정된 콘텐츠를 포함한다는 결정에 응답하여, 복수의 사운드 입력들 중 개별 사운드 입력들을 이용하여 기준 표현을 반복하여 조정하는 것을 포함한다.

[0022] 일부 구현예들에서, 이 방법은 전자 디바이스가 사전결정된 방향에 있는지 여부를 결정하는 것과, 전자 디바이스가 사전결정된 방향에 있다는 결정에 따라 음성 트리거의 사전결정된 모드를 활성화하는 것을 포함한다. 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 아래를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 스탠바이 모드이다. 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 위를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 청취 모드이다.

[0023] 일부 구현예들은 음성 트리거를 동작시키기 위한 방법을 제공한다. 이 방법은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다. 이 방법은 제1 모드에서 음성 트리거를 동작시키는 것을 포함한다. 이 방법은 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐색되었음(occluded)을 검출함으로써 전자 디바이스가 실질적으로 밀폐된 공간 내에 있는지 여부를 결정하는 것을 더 포함한다. 이 방법은 전자 디바이스가 실질적으로 밀폐된 공간 내에 있다는 결정에 따라, 음성 트리거를 제2 모드로 전환하는 것을 더 포함한다. 일부 구현예들에서, 제2 모드는 스탠바이 모드이다.

[0024] 일부 구현예들은 음성 트리거를 동작시키기 위한 방법을 제공한다. 이 방법은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다. 이 방법은 전자 디바이스가 사전결정된 방향에 있는지 여부를 결정하는 것과, 전자 디바이스가 사전결정된 방향에 있다는 결정에 따라 음성 트리거의 사전결정된 모드를 활성화하는 것을 포함한다. 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 아래를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 스탠바이 모드이다. 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 위를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 청취 모드이다.

[0025] 일부 구현예들에 따르면, 전자 디바이스는 사운드 입력을 수신하도록 구성된 사운드 수신 유닛; 및 사운드 수신 유닛에 연결된 프로세싱 유닛을 포함한다. 프로세싱 유닛은 사운드 입력의 적어도 일부가 사전결정된 타입의 사운드에 대응하는지 여부를 결정하고; 사운드 입력의 적어도 일부가 사전결정된 타입에 대응한다는 결정에 따라, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하며; 사운드 입력이 사전결정된 콘텐츠를 포함한다는 결정에 따라, 스피치-기반 서비스를 개시하도록 구성된다. 일부 구현예들에서, 프로세싱 유닛은 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하기 전에, 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하도록 추가로 구성된다. 일부 구현예들에서, 프로세싱 유닛은 사운드 입력이 특정 사용자의 음성에 대응하는지 여부를 결정하도록 추가로 구성된다.

[0026] 일부 구현예들에 따르면, 전자 디바이스는 복수의 모드들 중 제1 모드에서 음성 트리거를 동작시키도록 구성된 음성 트리거 유닛; 및 음성 트리거 유닛에 연결된 프로세싱 유닛을 포함한다. 일부 구현예들에서, 프로세싱 유닛은: 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐색되어 있음을 검출함으로써 전자 디바이스가 실질적으로 밀폐된 공간 내에 있는지 여부를 결정하고; 전자 디바이스가 실질적으로 밀폐된 공간 내에 있다는 결정에 따라, 음성 트리거를 제2 모드로 전환하도록 구성된다. 일부 구현예들에서, 프로세싱 유닛은 전자 디바이스가 사전결정된 방향에 있는지 여부를 결정하고; 전자 디바이스가 사전결정된 방향에 있다는 결정에 따라, 음성 트리거의 사전결정된 모드를 활성화하도록 구성된다.

[0027] 일부 구현예들에 따르면, 컴퓨터 판독가능한 저장 매체(예를 들어, 비일시적 컴퓨터 판독가능한 저장 매체)가 제공되며, 컴퓨터 판독가능한 저장 매체는 전자 디바이스의 하나 이상의 프로세서에 의해 실행하기 위한 하나 이상의 프로그램을 저장하고, 하나 이상의 프로그램은 본 명세서에 기술된 방법들 중 임의의 방법을 수행하기

위한 명령어들을 포함한다.

- [0028] 일부 구현예들에 따르면, 본 명세서에 기술된 방법들 중 임의의 방법을 수행하기 위한 수단을 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공된다.
- [0029] 일부 구현예들에 따르면, 본 명세서에 기술된 방법들 중 임의의 방법을 수행하도록 구성된 프로세싱 유닛을 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공된다.
- [0030] 일부 구현예들에 따르면, 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 하나 이상의 프로그램을 저장하는 메모리를 포함하는 전자 디바이스(예를 들어, 휴대용 전자 디바이스)가 제공되며, 하나 이상의 프로그램은 본 명세서에 기술된 방법들 중 임의의 방법을 수행하기 위한 명령어들을 포함한다.
- [0031] 일부 구현예들에 따르면, 전자 디바이스에서 사용하기 위한 정보 프로세싱 장치가 제공되며, 이러한 정보 프로세싱 장치는 본 명세서에 기술된 방법들 중 임의의 방법을 수행하기 위한 수단을 포함한다.

도면의 간단한 설명

- [0032] 도 1은 일부 구현예들에 따른 디지털 어시스턴트가 동작하는 환경을 도시하는 블록 다이어그램이다.
 - 도 2는 일부 구현예들에 따른 디지털 어시스턴트 클라이언트 시스템을 도시하는 블록 다이어그램이다.
 - 도 3a는 일부 구현예들에 따른 독립형 디지털 어시스턴트 시스템 또는 디지털 어시스턴트 서버 시스템을 도시하는 블록 다이어그램이다.
 - 도 3b는 일부 구현예들에 따른 도 3a에 도시된 디지털 어시스턴트의 기능들을 도시하는 블록 다이어그램이다.
 - 도 3c는 일부 구현예들에 따른 온톨로지(ontology)의 일부분을 도시하는 네트워크도이다.
 - 도 4는 일부 구현예들에 따른 음성 트리거 시스템의 컴포넌트들을 도시한 블록 다이어그램이다.
 - 도 5 내지 도 7은 일부 구현예들에 따른 음성 트리거 시스템을 동작하기 위한 방법들을 도시한 순서도이다.
 - 도 8 및 도 9는 일부 구현예들에 따른 전자 디바이스의 기능 블록 다이어그램이다.
- 동일한 도면 부호들은 도면들 전체에 걸쳐 대응하는 부분들을 가리킨다.

발명을 실시하기 위한 구체적인 내용

- [0033] 도 1은 일부 구현예들에 따른 디지털 어시스턴트의 동작 환경(100)의 블록 다이어그램이다. "디지털 어시스턴트", "가상 어시스턴트", "인텔리전트 자동화 어시스턴트", "음성-기반 디지털 어시스턴트", 또는 "자동 디지털 어시스턴트"와 같은 용어들은 사용자 의도를 추론하기 위해서 자연 언어 입력을 구두의(spoken) 및/또는 텍스트의(textual) 형태로 해석하고(예를 들어, 자연 언어 입력에 상응하는 작업 타입을 식별), 추론된 사용자 의도에 기초하여 행동들을 수행하는(예를 들어, 식별된 작업 타입에 상응하는 작업을 수행) 임의의 정보 프로세싱 시스템을 지칭한다. 예를 들면, 추론된 사용자 의도를 좇아 행동하기 위해, 시스템은 다음의 것 중 하나 이상을 수행할 수 있다: 추론된 사용자 의도를 달성하도록 설계된 단계들 및 파라미터들을 갖는 작업 흐름을 식별하는 단계(예를 들어, 작업 유형을 식별하는 단계), 추론된 사용자 의도로부터의 특정 요구사항들을 작업 흐름에 입력하는 단계, 프로그램, 방법, 서비스, API 등을 호출함으로써 작업 흐름을 실행하는 단계(예를 들어, 서비스 제공자에게 요청을 송신하는 단계), 및 청각적(예를 들어, 스피치) 및/또는 시각적 형태로 사용자에게 대한 출력 응답을 생성하는 단계.
- [0034] 특히, 한번 개시되면, 디지털 어시스턴트 시스템은 적어도 부분적으로 자연 언어 커맨드, 요청, 진술, 서술 및/또는 조회의 형태인 사용자 요청을 수용할 수 있다. 통상적으로, 사용자 요청은 디지털 어시스턴트 시스템을 통해 정보를 제공하는 대답 또는 작업의 수행 중 하나를 요구한다. 사용자 요청에 대한 만족스러운 응답은 일반적으로 요구한 정보를 제공하는 대답의 제공, 요구한 작업의 수행 중 하나, 또는 이 둘의 조합이다. 예를 들어, 사용자는 디지털 어시스턴트 시스템에게 "지금 내가 어디에 있습니까?"와 같은 질문을 할 수 있다. 사용자의 현재 위치에 기초하여, 디지털 어시스턴트는 "당신은 서문 근처 센트럴파크에 있습니다"라고 대답할 수 있다. 사용자는 또한 예를 들어 "나의 친구들을 다음주 내 여자친구의 생일파티에 초대해주세요"라고 진술함으로써 작업의 수행을 요청할 수 있다. 그에 대한 응답으로, 디지털 어시스턴트는 "네, 바로 실행하겠습니다"라는 음성 출력을 생성함으로써 요청을 수신하였음을 알린 후에 사용자의 이메일 주소로부터 사용자의 전자 주소록 또는 연락처 목록에 나열된 사용자의 친구들 각각에게 적절한 캘린더 초대를 전송할 수 있다. 정보 또는 다

양한 작업들의 수행을 요청하기 위해 디지털 어시스턴트와 상호작용하는 많은 다른 방식들이 존재한다. 언어적 응답을 제공하는 것과 프로그래밍된 행동을 취하는 것 외에도, 디지털 어시스턴트는 또한 (예를 들어, 텍스트, 알람, 음악, 비디오, 애니메이션 등과 같은) 다른 시각적 또는 청각적 형태의 응답도 제공할 수 있다.

[0035] 도 1에 도시된 바와 같이, 일부 구현예들에서, 디지털 어시스턴트 시스템은 클라이언트-서버 모델에 따라 구현된다. 디지털 어시스턴트 시스템은 사용자 디바이스(예를 들어, (104a, 104b)) 상에서 실행되는 클라이언트 측 부분(예를 들어, (102a, 102b))(아래에서는 "디지털 어시스턴트(DA) 클라이언트(102)") 및 서버 시스템(108) 상에서 실행되는 서버 측 부분(106)(아래에서는 "디지털 어시스턴트(DA) 서버(106)")를 포함한다. DA 클라이언트(102)는 하나 이상의 네트워크(110)를 통해 DA 서버(106)와 통신한다. DA 클라이언트(102)는, 사용자 대면(user-facing) 입출력 프로세싱 및 DA 서버(106)와의 통신과 같은 클라이언트 측 기능들을 제공한다. DA 서버(106)는 (클라이언트 디바이스 또는 전자 디바이스로도 지칭되는) 개별 사용자 디바이스(104) 상에 각각 존재하는 임의의 수의 DA 클라이언트(102)에 대한 서버 측 기능들을 제공한다.

[0036] 일부 구현예들에서, DA 서버(106)는 클라이언트-대면 I/O 인터페이스(112), 하나 이상의 프로세싱 모듈(114), 데이터 및 모델(116), 외부 서비스에 대한 I/O 인터페이스(118), 사진 및 태그 데이터베이스(130) 및 사진-태그 모듈(132)을 포함한다. 클라이언트-대면 I/O 인터페이스는 디지털 어시스턴트 서버(106)에 대한 클라이언트-대면 입출력 프로세싱을 가능하게 한다. 하나 이상의 프로세싱 모듈(114)은 데이터 및 모델(116)을 활용하여 자연 언어 입력에 기초해 사용자의 의도를 결정하고 추론된 사용자 의도에 기초하여 작업 실행을 수행한다. 사진 및 태그 데이터베이스(130)는 디지털 사진들의 팽거프린트 및 선택적으로는 디지털 사진들 자체뿐만 아니라, 디지털 사진들과 연관된 태그들을 저장한다. 사진-태그 모듈(132)은 태그를 생성하고, 사진 및/또는 팽거프린트와 연관하여 태그를 저장하고, 사진을 자동으로 태그하고, 태그를 사진 내의 위치에 링크시킨다.

[0037] 일부 구현예들에서, DA 서버(106)는 작업 완료 또는 정보 획득을 위해 네트워크(들)(110)를 통해 외부 서비스(120)(예를 들어, 내비게이션 서비스(들)(122-1), 메시징 서비스(들)(122-2), 정보 서비스(들)(122-3), 캘린더 서비스(122-4), 텔레포니 서비스(122-5), 사진 서비스(들)(122-6) 등)와 통신한다. 외부 서비스(118)에 대한 I/O 인터페이스는 이러한 통신을 용이하게 한다.

[0038] 사용자 디바이스(104)의 예에는 핸드헬드 컴퓨터, 개인용 디지털 어시스턴트(PDA), 태블릿 컴퓨터, 랩톱 컴퓨터, 데스크탑 컴퓨터, 셀룰러 전화기, 스마트폰, 개선된 일반 패킷 무선 서비스(enhanced general packet radio service, EGPRS) 모바일 전화기, 미디어 플레이어, 탐색 디바이스, 게임 콘솔, 텔레비전, 원격 조종기, 또는 이러한 데이터 프로세싱 디바이스들 중 임의의 둘 이상의 조합 또는 임의의 다른 적절한 데이터 프로세싱 디바이스들이 포함되지만, 이에 제한되지는 않는다. 사용자 디바이스(104)에 대한 더 자세한 내용은 도 2에 도시된 예시적인 사용자 디바이스(104)를 참조하여 제공된다.

[0039] 통신 네트워크(들)(110)의 예에는 로컬 영역 네트워크(LAN) 및 예로서 인터넷과 같은 광역 네트워크(WAN)가 포함된다. 통신 네트워크(들)(110)는 다양한 유선 또는 무선 프로토콜, 예컨대 이더넷(Ethernet), 범용 직렬 버스(Universal Serial Bus, USB), 화이어와이어(FIREWIRE), 모바일 통신용 글로벌 시스템(Global System for Mobile Communication, GSM), 개선된 데이터 GSM 환경(Enhanced Data GSM Environment, EDGE), 코드 분할 다중 접속(code division multiple access, CDMA), 시간 분할 다중 접속(time division multiple access, TDMA), 블루투스(Bluetooth), Wi-Fi, VoIP(voice over Internet Protocol), Wi-MAX, 또는 임의의 다른 적합한 통신 프로토콜을 포함하는 임의의 알려진 네트워크 프로토콜을 사용하여 구현될 수 있다.

[0040] 서버 시스템(108)은 적어도 하나의 데이터 프로세싱 장치 및/또는 분산된 컴퓨터들의 네트워크 상에서 구현될 수 있다. 일부 구현예들에서, 서버 시스템(108)은 또한 제3자 서비스 제공자들(예컨대, 제3자 클라우드 서비스 제공자들)의 다양한 가상 디바이스들 및/또는 서비스들을 사용하여, 서버 시스템(108)의 기초 컴퓨팅 리소스들 및/또는 인프라구조 리소스들을 제공한다.

[0041] 도 1에 도시된 디지털 어시스턴트 시스템이 클라이언트 측 부분(예를 들어, DA 클라이언트(102)) 및 서버 측 부분(예를 들어, DA 서버(106)) 모두를 포함하지만, 일부 구현예들에서, 디지털 어시스턴트 시스템은 오직 서버 측 부분(예를 들어 DA 서버(106))만을 지칭한다. 일부 구현예들에서, 디지털 어시스턴트의 기능들은 사용자 디바이스 상에 설치된 독립형 애플리케이션으로서 구현될 수 있다. 또한, 디지털 어시스턴트의 클라이언트 부분과 서버 부분 사이의 기능들의 분담은 상이한 구현예들에서 다를 수 있다. 예를 들어, 일부 구현예들에서, DA 클라이언트(102)는 오직 사용자-대면 입력 및 출력 프로세싱 기능들만을 제공하고 디지털 어시스턴트의 다른 모든 기능들을 DA 서버(106)에 위임하는 쉘-클라이언트(thin-client)이다. 일부 다른 구현예들에서, DA 클라이언트(102)는 DA 서버(106)의 하나 이상의 기능들을 수행하거나 보조하도록 구성된다.

- [0042] 도 2는 일부 구현예들에 따른 사용자 디바이스(104)의 블록 다이어그램이다. 사용자 디바이스(104)는 메모리 인터페이스(202), 하나 이상의 프로세서(204) 및 주변장치 인터페이스(206)를 포함한다. 사용자 디바이스(104)의 다양한 컴포넌트들은 하나 이상의 통신 버스들 또는 신호 라인들에 의해 연결된다. 사용자 디바이스(104)는 주변장치 인터페이스(206)에 연결된 다양한 센서들, 서브시스템들 및 주변 디바이스들을 포함한다. 센서들, 서브시스템들 및 주변 디바이스들은 정보를 수집하고/하거나 사용자 디바이스(104)의 다양한 기능들을 가능하게 한다.
- [0043] 예를 들어, 일부 구현예들에서, 동작 센서(210)(예를 들어, 가속도계), 광 센서(212), GPS 수신기(213), 온도 센서 및 근접 센서(214)가 방향, 광 및 근접 감지 기능을 용이하게 하도록 주변장치 인터페이스(206)에 연결된다. 일부 구현예들에서, 생체 인식 센서, 기압계 등과 같은 다른 센서들(216)이 주변장치 인터페이스(206)에 접속되어 관련 기능들을 가능하게 한다.
- [0044] 일부 구현예들에서, 사용자 디바이스(104)는 주변장치 인터페이스(206)에 연결된 카메라 서브시스템(220)을 포함한다. 일부 구현예들에서, 카메라 서브시스템(220)의 광학 센서(222)는 사진 촬영 및 비디오 클립 녹화와 같은 카메라 기능들을 가능하게 한다. 일부 구현예들에서, 사용자 디바이스(104)는 통신 기능들을 제공하는 하나 이상의 유선 및/또는 무선 통신 서브시스템(224)을 포함한다. 통신 서브시스템(224)은 통상적으로 다양한 통신 포트, 무선 주파수 수신기 및 송신기 및/또는 광학(예를 들어, 적외선) 수신기 및 송신기를 포함한다. 일부 구현예들에서, 사용자 디바이스(104)는 하나 이상의 스피커(228) 및 하나 이상의 마이크로폰(230)에 연결된 오디오 서브시스템(226)을 포함하여, 음성 인식, 음성 복제, 디지털 녹음 및 전화 기능들과 같은 음성-인에이블형 기능들을 가능하게 한다. 일부 구현예들에서, 오디오 서브시스템(226)은 음성 트리거 시스템(400)에 연결된다. 일부 구현예들에서, 음성 트리거 시스템(400) 및/또는 오디오 서브시스템(226)은 사운드 입력을 수신 및/또는 분석하기 위한 저전력 오디오 회로 및/또는 프로그램(즉, 하드웨어 및/또는 소프트웨어를 포함함)을 포함하며, 이들은 예를 들어 하나 이상의 아날로그-디지털 변환기, 디지털 신호 프로세서(DSP), 사운드 검출기, 메모리 버퍼, 코덱 등을 포함한다. 일부 구현예들에서, 저전력 오디오 회로는 (단독으로 또는 사용자 디바이스(104)의 다른 컴포넌트들에 더하여) 음성-기반 디지털 어시스턴트 또는 다른 스피치-기반 서비스와 같은 사용자 디바이스(104)의 하나 이상의 양태에 대한 음성(또는 사운드) 트리거 기능을 제공한다. 일부 구현예들에서, 저전력 오디오 회로는 프로세서(들)(204), I/O 서브시스템(240), 메모리(250) 등과 같은 사용자 디바이스(104)의 다른 컴포넌트들이 쉼다운되고/되거나 스탠바이 모드에 있을 때에도 음성 트리거 기능을 제공한다. 음성 트리거 시스템(400)은 도 4와 관련하여 더욱 자세하게 기술된다.
- [0045] 일부 구현예들에서, I/O 서브시스템(240)이 또한 주변장치 인터페이스(206)에 연결된다. 일부 구현예들에서, 사용자 디바이스(104)는 터치 스크린(246)을 포함하고, I/O 서브시스템(240)은 터치 스크린(246)에 연결된 터치 스크린 제어기(242)를 포함한다. 사용자 디바이스(104)가 터치 스크린(246) 및 터치 스크린 제어기(242)를 포함할 때, 터치 스크린(246) 및 터치 스크린 제어기(242)는 통상적으로, 예를 들어 용량성, 저항성, 적외선, 표면 음파 기술, 근접 센서 어레이 등과 같은 복수의 터치 감지 기술들 중 임의의 기술을 이용하여 터치 스크린의 접촉 및 이동 또는 고장을 검출하도록 구성된다. 일부 구현예들에서, 사용자 디바이스(104)는 터치 감지 표면을 포함하지 않는 디스플레이를 포함한다. 일부 구현예들에서, 사용자 디바이스(104)는 별개의 터치 감지 표면을 포함한다. 일부 구현예들에서, 사용자 디바이스(104)는 다른 입력 제어기(들)(244)를 포함한다. 사용자 디바이스(104)가 다른 입력 제어기(들)(244)를 포함하는 경우, 다른 입력 제어기(들)(244)는 통상적으로 다른 입력/제어 디바이스(248), 예컨대 하나 이상의 버튼, 로커 스위치(rocker switch), 썸 휠(thumb-wheel), 적외선 포트, USB 포트 및/또는 스타일러스와 같은 포인터 디바이스에 연결된다.
- [0046] 메모리 인터페이스(202)는 메모리(250)에 연결된다. 일부 구현예들에서, 메모리(250)는 고속 랜덤 액세스 메모리 및/또는 비휘발성 메모리와 같은 비일시적(non-transitory) 컴퓨터 판독가능한 매체를 포함한다(예를 들어, 하나 이상의 자기 디스크 저장 디바이스, 하나 이상의 플래시 메모리 디바이스, 하나 이상의 광학 저장 디바이스 및/또는 다른 비휘발성 고체 상태 메모리 디바이스). 일부 구현예들에서, 메모리(250)는 운영 시스템(252), 통신 모듈(254), 그래픽 사용자 인터페이스 모듈(256), 센서 프로세싱 모듈(258), 폰 모듈(260), 애플리케이션(262) 및 이들의 하위세트 또는 상위세트를 저장한다. 운영 시스템(252)은 기본 시스템 서비스를 핸들링하고 하드웨어 의존 작업들을 수행하기 위한 명령어들을 포함한다. 통신 모듈(254)은 하나 이상의 추가 디바이스, 하나 이상의 컴퓨터 및/또는 하나 이상의 서버와의 통신을 가능하게 한다. 그래픽 사용자 인터페이스 모듈(256)은 그래픽 사용자 인터페이스 프로세싱을 가능하게 한다. 센서 프로세싱 모듈(258)은 센서 관련 프로세싱 및 기능(예를 들어, 하나 이상의 마이크로폰(228)을 이용하여 수신된 음성 입력을 프로세싱함)을 가능하게 한다. 전화 모듈(260)은 전화 관련 프로세스 및 기능들을 가능하게 한다. 애플리케이션 모듈(262)은, 전자 메시

징, 웹브라우저, 미디어 프로세싱, 내비게이션, 이미징 및/또는 다른 프로세스들 및 기능들과 같은 사용자 애플리케이션들의 다양한 기능성을 가능하게 한다. 일부 구현예들에서, 사용자 디바이스(104)는 외부 서비스 제공자들 중 적어도 하나와 각각 연관된 하나 이상의 소프트웨어 애플리케이션(270-1, 270-2)을 메모리(250)에 저장한다.

[0047] 전술된 바와 같이, 일부 구현예들에서, 메모리(250)는 또한 디지털 어시스턴트의 클라이언트 측 기능들을 제공하기 위해 (예컨대, 디지털 어시스턴트 클라이언트 모듈(264) 내의) 클라이언트 측 디지털 어시스턴트 명령어들 및 다양한 사용자 데이터(266)(예컨대, 사용자-특정 어휘 데이터, 선호도 데이터 및/또는 사용자의 전자 주소록 또는 연락처 목록, 할 일 목록, 쇼핑 목록 등과 같은 다른 데이터)를 저장한다.

[0048] 다양한 구현예들에서, 디지털 어시스턴트 클라이언트 모듈(264)은 사용자 디바이스(104)의 다양한 사용자 인터페이스(예컨대, I/O 서브시스템(244))를 통해 음성 입력, 텍스트 입력, 터치 입력 및/또는 제스처 입력을 수할 수 있다. 디지털 어시스턴트 클라이언트 모듈(264)은 또한 청각적, 시각적 및/또는 촉각적 형태의 출력을 제공하는 것이 가능하다. 예를 들면, 출력은 음성, 사운드, 알람, 텍스트 메시지, 메뉴, 그래픽, 비디오, 애니메이션, 진동 및/또는 상기한 것들 중 2개 이상의 조합으로서 제공될 수 있다. 작동 중에, 디지털 어시스턴트 클라이언트 모듈(264)은 통신 서브시스템(224)을 이용하여 디지털 어시스턴트 서버(예를 들어, 디지털 어시스턴트 서버(106), 도 1)와 통신한다.

[0049] 일부 구현예들에서, 디지털 어시스턴트 클라이언트 모듈(264)은 사용자 입력과 연관된 콘텍스트(context)를 확립하기 위해 다양한 센서, 서브시스템 및 주변장치 디바이스를 활용하여 사용자 디바이스(104)의 주변 환경으로부터 추가 정보를 수집한다. 일부 구현예들에서, 디지털 어시스턴트 클라이언트 모듈(264)은 사용자 입력과 함께 콘텍스트 정보 또는 그의 하위세트를 디지털 어시스턴트 서버(예컨대, 디지털 어시스턴트 서버(106), 도 1)에 제공하여 사용자의 의도를 추론하는 것을 돕는다.

[0050] 일부 구현예들에서, 사용자 입력에 동반될 수 있는 콘텍스트 정보는 센서 정보, 예컨대 조명, 주변 소음, 주변 온도, 주위 환경의 이미지 또는 비디오 등을 포함한다. 일부 구현예들에서, 콘텍스트 정보는 또한 디바이스의 물리적 상태, 예컨대 디바이스 방향, 디바이스 위치, 디바이스 온도, 전력 레벨, 속도, 가속도, 동작 패턴들, 셀룰러 신호 강도 등을 포함한다. 일부 구현예들에서, 사용자 디바이스(106)의 소프트웨어 상태에 관련된 정보, 예컨대 사용자 디바이스(104)의 실행 중인 프로세스들, 설치된 프로그램들, 과거 및 현재의 네트워크 활동성들, 백그라운드 서비스들, 에러 로그들, 리소스 사용 등이 또한 사용자 입력과 연관된 콘텍스트 정보로서 디지털 어시스턴트 서버(예컨대, 디지털 어시스턴트 서버(106), 도 1)에 제공된다.

[0051] 일부 구현예들에서, DA 클라이언트 모듈(264)은 디지털 어시스턴트 서버로부터의 요청들에 응답하여 사용자 디바이스(104) 상에 저장된 정보(예컨대, 사용자 데이터(266)의 적어도 일부분)를 선택적으로 제공한다. 일부 구현예들에서, 디지털 어시스턴트 클라이언트 모듈(264)은 또한 디지털 어시스턴트 서버(106)(도 1)에 의한 요청시에 자연 언어 대화 또는 다른 사용자 인터페이스들을 통해 사용자로부터 추가 입력을 이끌어낸다. 디지털 어시스턴트 클라이언트 모듈(264)은, 사용자 요청에서 표현된 사용자의 의도의 의도 추론 및/또는 이행에서 디지털 어시스턴트 서버(106)를 돕기 위해, 추가적인 입력을 디지털 어시스턴트 서버(106)에 전달한다.

[0052] 일부 구현예들에서, 메모리(250)는 추가 명령어들을 포함할 수 있거나 또는 더 적은 명령어들을 포함할 수 있다. 또한, 사용자 디바이스(104)의 다양한 기능들이 하나 이상의 신호 프로세싱 및/또는 애플리케이션 특정 집적 회로에 포함되는 하드웨어 및/또는 펌웨어에서 구현될 수 있으며, 따라서 사용자 디바이스(104)가 도 2에 도시된 모든 모듈 및 애플리케이션을 포함할 필요는 없다.

[0053] 도 3a는 일부 구현예들에 따른 예시적인 디지털 어시스턴트 시스템(300)(디지털 어시스턴트로도 지칭됨)의 블록 다이어그램이다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 독립형 컴퓨터 시스템 상에서 구현된다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 다수의 컴퓨터들에 걸쳐서 분포된다. 일부 구현예들에서, 디지털 어시스턴트의 모듈들 및 기능들 중 일부는 서버 부분과 클라이언트 부분으로 나뉘는데, 여기서 클라이언트 부분은 사용자 디바이스(예컨대, 사용자 디바이스(104)) 상에 존재하고, 예컨대 도 1에 도시된 바와 같은 하나 이상의 네트워크들을 통해 서버 부분(예컨대, 서버 시스템(108))과 통신한다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 도 1에 도시된 서버 시스템(108)(및/또는 디지털 어시스턴트 서버(106))의 실시예이다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 사용자 디바이스(예컨대, 사용자 디바이스(104), 도 1)에서 구현되며, 그에 따라 클라이언트-서버 시스템에 대한 필요성을 제거한다. 디지털 어시스턴트 시스템(300)은 디지털 어시스턴트 시스템의 일 예에 불과하며, 디지털 어시스턴트 시스템(300)은 도시된 것보다 더 많은 또는 더 적은 컴포넌트들을 구비할 수도 있거나, 2개 이상의 컴포넌트들을 병합할 수도 있거나

나, 또는 컴포넌트들의 상이한 구성 또는 배치를 가질 수도 있음을 주목해야만 한다. 도 3a에 도시된 다양한 컴포넌트들은 하나 이상의 신호 프로세싱 및/또는 애플리케이션 특정 집적 회로를 포함하는 하드웨어, 소프트웨어, 펌웨어, 또는 이들의 조합에서 구현될 수 있다.

[0054] 디지털 어시스턴트 시스템(300)은 메모리(302), 하나 이상의 프로세서(304), 입력/출력(I/O) 인터페이스(306) 및 네트워크 통신 인터페이스(308)를 포함한다. 이러한 컴포넌트들은 하나 이상의 통신 버스 또는 신호 라인(310)을 통해 서로 통신한다.

[0055] 일부 구현예들에서, 메모리(302)는 고속 랜덤 액세스 메모리 및/또는 비휘발성 컴퓨터 판독가능 저장 매체와 같은 비일시적 컴퓨터 판독가능 매체를 포함한다(예컨대, 하나 이상의 자기 디스크 저장 디바이스, 하나 이상의 플래시 메모리 디바이스, 하나 이상의 광학 저장 디바이스 및/또는 다른 비휘발성 고체 상태 메모리 디바이스들).

[0056] I/O 인터페이스(306)는 디스플레이, 키보드, 터치 스크린 및 마이크로폰과 같은 디지털 어시스턴트 시스템(300)의 입력/출력 디바이스들(316)을 사용자 인터페이스 모듈(322)에 연결한다. I/O 인터페이스(306)는, 사용자 인터페이스 모듈(322)과 함께, 사용자 입력(예를 들어, 음성 입력, 키보드 입력, 터치 입력 등)을 수신하여 그에 따라 그것들을 프로세싱한다. 일부 구현예들에서, 디지털 어시스턴트가 독립형 사용자 디바이스 상에서 구현되는 경우, 디지털 어시스턴트 시스템(300)은 도 2의 사용자 디바이스(104)에 대해 기술된 컴포넌트들 및 I/O 및 통신 인터페이스들 중 임의의 것(예컨대, 하나 이상의 마이크로폰들(230))을 포함한다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 디지털 어시스턴트 구현예의 서버 부분을 나타내고, 사용자 디바이스(예컨대, 도 2에 도시된 사용자 디바이스(104)) 상에 존재하는 클라이언트 측 부분을 통해 사용자와 상호작용한다.

[0057] 일부 구현예들에서, 네트워크 통신 인터페이스(308)는 유선 통신 포트(들)(312) 및/또는 무선 송신 및 수신 회로(314)를 포함한다. 유선 통신 포트(들)는 하나 이상의 유선 인터페이스, 예를 들어, 이더넷, 범용 직렬 버스(USB), 파이어와이어 등을 통해 통신 신호를 수신하고 송신한다. 무선 회로(314)는 통상적으로 통신 네트워크 및 기타 통신 디바이스로/로부터 RF 신호 및/또는 광학 신호를 수신하고 송신한다. 무선 통신들은, GSM, EDGE, CDMA, TDMA, 블루투스, 와이파이, VoIP, 와이맥스, 또는 임의의 다른 적절한 통신 프로토콜과 같은 복수의 통신 표준들, 프로토콜들 및 기술들 중 임의의 것을 사용할 수도 있다. 네트워크 통신 인터페이스(308)는 인터넷, 인트라넷 및/또는 무선 네트워크, 예컨대 셀룰러 전화 네트워크, 무선 근거리 통신망(LAN) 및/또는 도시권 통신망(metropolitan area network, MAN)과 같은 네트워크들을 통해 디지털 어시스턴트 시스템(300)과 다른 디바이스들 사이의 통신을 가능하게 한다.

[0058] 일부 구현예들에서, 메모리(302)의 비일시적 컴퓨터 판독가능 저장 매체는: 프로그램, 모듈, 명령어 및 운영 시스템(318), 통신 모듈(320), 사용자 인터페이스 모듈(322), 하나 이상의 애플리케이션(324) 및 디지털 어시스턴트 모듈(326) 모두 또는 이들의 하위세트를 포함하는 데이터 구조를 저장한다. 하나 이상의 프로세서(304)는 이들 프로그램, 모듈 및 명령어를 실행하고, 데이터 구조로부터/로 판독/기록한다.

[0059] 운영 시스템(318)(예컨대, 다윈(Darwin), RTXC, LINUX, UNIX, OS X, iOS, WINDOWS, 또는 VxWorks와 같은 내장형 운영 시스템)은 일반적인 시스템 작업들(예컨대, 메모리 관리, 저장 디바이스 제어, 전력 관리 등)을 제어하고 관리하기 위한 다양한 소프트웨어 컴포넌트들 및/또는 드라이버들을 포함하고, 다양한 하드웨어, 펌웨어 및 소프트웨어 컴포넌트들 간의 통신을 가능하게 한다.

[0060] 통신 모듈(320)은 네트워크 통신 인터페이스(308)를 통해 다른 디바이스들과 디지털 어시스턴트 시스템(300) 사이의 통신을 가능하게 한다. 예를 들어, 통신 모듈(320)은 도 2에 도시된 디바이스(104)의 통신 모듈(254)과 통신할 수 있다. 통신 모듈(320)은 또한 무선 회로(314) 및/또는 유선 통신 포트(312)에 의해 수신되는 데이터를 핸들링하기 위한 다양한 소프트웨어 컴포넌트들을 포함한다.

[0061] 일부 구현예들에서, 사용자 인터페이스 모듈(322)은 I/O 인터페이스(306)를 통해 사용자로부터(예컨대, 키보드, 터치 스크린 및/또는 마이크로폰으로부터) 커맨드 및/또는 입력을 수신하고, 디스플레이 상에 사용자 인터페이스 객체들을 제공한다.

[0062] 애플리케이션들(324)은, 하나 이상의 프로세서(304)에 의해 실행되도록 구성된 프로그램들 및/또는 모듈들을 포함한다. 예를 들어, 디지털 어시스턴트 시스템이 독립형 사용자 디바이스 상에서 구현되면, 애플리케이션들(324)은, 게임, 캘린더 애플리케이션, 내비게이션 애플리케이션, 또는 이메일 애플리케이션과 같은 사용자 애플리케이션들을 포함할 수도 있다. 디지털 어시스턴트 시스템(300)이 서버 팜(server farm) 상에 구현되면, 애플리케이션들(324)은, 예를 들어, 리소스 관리 애플리케이션, 진단 애플리케이션, 또는 스케줄링 애플리케이션을

포함할 수도 있다.

[0063] 메모리(302)는 또한 디지털 어시스턴트 모듈(또는 디지털 어시스턴트의 서버 부분)(326)을 저장한다. 일부 구현예들에서, 디지털 어시스턴트 모듈(326)은: 입력/출력 프로세싱 모듈(328), 스피치-텍스트(STT) 프로세싱 모듈(330), 자연 언어 프로세싱 모듈(332), 대화 흐름 프로세싱 모듈(334), 작업 흐름 프로세싱 모듈(336), 서비스 프로세싱 모듈(338) 및 사진 모듈(132)과 같은 서브 모듈들, 또는 이들의 하위세트 또는 상위세트를 포함한다. 이 프로세싱 모듈들의 각각은: 온톨로지(360), 어휘 인덱스(344), 사용자 데이터(348), 분류 모듈(349), 중의성 해결(disambiguation) 모듈(350), 작업 흐름 모델(354), 서비스 모델(356), 사진 태깅 모듈(358), 탐색 모듈(360) 및 로컬 태그/사진 스토리지(362)와 같은 디지털 어시스턴트(326)의 데이터 및 모델 중 하나 이상, 또는 이들의 하위세트 또는 상위세트에 대한 액세스를 갖는다.

[0064] 일부 구현예들에서, 디지털 어시스턴트 모듈(326)에서 구현되는 프로세싱 모듈들(예컨대, 입력/출력 프로세싱 모듈(328), STT 프로세싱 모듈(330), 자연 언어 프로세싱 모듈(332), 대화 흐름 프로세싱 모듈(334), 작업 흐름 프로세싱 모듈(336) 및/또는 서비스 프로세싱 모듈(338)), 데이터 및 모델들을 이용하여, 디지털 어시스턴트 시스템(300)은: 사용자로부터 수신된 자연 언어 입력에서 표현된 사용자의 의도를 식별하는 것; 사용자의 의도를 완전히 추론하는 데 필요한 정보를 (예를 들어, 단어, 이름, 의도 등의 중의성을 해결함으로써) 능동적으로 유도하고 획득하는 것; 추론된 의도를 이행하기 위한 작업 흐름을 결정하는 것; 및 추론된 의도를 이행하기 위해 작업 흐름을 실행하는 것 중 적어도 일부를 수행한다. 일부 구현예들에서, 디지털 어시스턴트는 또한 다양한 이유들로 만족스러운 응답이 사용자에게 제공되지 않았거나 제공될 수 없었을 때 적절한 행동들을 취한다.

[0065] 일부 구현예들에서, 아래에서 논의되는 바와 같이, 디지털 어시스턴트 시스템(300)이 자연 언어 입력으로부터 디지털 사진에 태그하라는 사용자의 의도를 식별하고, 적절한 정보를 이용하여 디지털 사진을 태그하도록 자연 언어 입력을 프로세싱한다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 자연 언어 입력을 이용한 디지털 사진 탐색, 사진의 자동-태그 등과 같이 사진과 관련된 다른 작업들도 수행한다. 도 3b에 도시된 바와 같이, 일부 구현예들에서, I/O 프로세싱 모듈(328)은 도 3a의 I/O 디바이스(316)를 통해 사용자와 상호작용하거나 또는 도 3a의 네트워크 통신 인터페이스(308)를 통해 사용자 디바이스(예를 들어, 도 1의 사용자 디바이스(104))와 상호작용하여 사용자 입력(예로서, 스피치 입력)을 획득하고 사용자 입력에 대한 응답을 제공한다. I/O 프로세싱 모듈(328)은, 사용자 입력의 수신과 함께 또는 사용자 입력의 수신 직후에, 사용자 디바이스로부터 사용자 입력과 관련된 콘텍스트 정보를 선택적으로 획득한다. 콘텍스트 정보(context information)는 사용자 입력과 관련이 있는 사용자 특정 데이터, 어휘 및/또는 선호사항을 포함한다. 일부 구현예들에서, 콘텍스트 정보는 또한 사용자 요청이 수신될 때의 디바이스(예컨대, 도 1의 사용자 디바이스(104))의 소프트웨어 및 하드웨어 상태들 및/또는 사용자 요청이 수신될 때의 사용자의 주변 환경에 관련된 정보를 포함한다. 일부 구현예들에서, I/O 프로세싱 모듈(328)은 또한 사용자 요청에 관하여 사용자에게 후속 질문들을 전송하고, 그로부터 답변들을 수신한다. 일부 구현예들에서, 사용자 요청이 I/O 프로세싱 모듈(328)에 의해 수신되고 사용자 요청이 스피치 입력을 포함하는 경우, I/O 프로세싱 모듈(328)은 스피치-텍스트 변환을 위해 스피치 입력을 스피치-텍스트(STT) 프로세싱 모듈(330)로 보낸다.

[0066] 일부 구현예들에서, 스피치-텍스트 프로세싱 모듈(330)은 I/O 프로세싱 모듈(328)을 통해 스피치 입력(예컨대, 음성 녹음에서 캡처된 사용자 발언)을 수신한다. 일부 구현예들에서, 스피치-텍스트 프로세싱 모듈(330)은 음소들의 시퀀스 및 궁극적으로, 하나 이상의 언어들로 기록되는 단어들 또는 토큰(token)들의 시퀀스로서 스피치 입력을 인식하기 위해 다양한 음향 및 언어 모델들을 이용한다. 스피치-텍스트 프로세싱 모듈(330)은 임의의 적절한 스피치 인식 기술, 음향 모델 및 언어 모델, 예컨대 히든 마르코프 모델(Hidden Markov Model), 동적 타임 워핑(Dynamic Time Warping, DTW) 기반 스피치 인식 및 기타 통계 및/또는 분석 기술을 사용하여 구현된다. 일부 구현예들에서, 스피치-텍스트 프로세싱은 적어도 부분적으로 제3자 서비스에 의해 또는 사용자의 디바이스 상에서 수행될 수 있다. 스피치-텍스트 프로세싱 모듈(330)이 스피치-텍스트 프로세싱의 결과물(예를 들어, 단어들 또는 토큰들의 시퀀스)을 획득하면, 의도 추론을 위해 결과물을 자연 언어 프로세싱 모듈(332)로 전달한다. 디지털 어시스턴트(326)의 자연 언어 프로세싱 모듈(332)("자연 언어 프로세서")은 스피치-텍스트 프로세싱 모듈(330)에 의해 생성된 단어들 또는 토큰들의 시퀀스("토큰 시퀀스")를 취하여, 이러한 토큰 시퀀스를 디지털 어시스턴트에 의해 인식된 하나 이상의 "행동가능한 의도(actionable intent)"와 연관시키고자 시도한다. 본 명세서에서 사용되는 "행동가능한 의도"는 디지털 어시스턴트(326) 및/또는 디지털 어시스턴트 시스템(300)(도 3a)에 의해 수행될 수 있는 작업을 나타내며, 작업 흐름 모델(354)에서 구현되는 연관된 작업 흐름을 갖는다. 연관된 작업 흐름은 디지털 어시스턴트 시스템(300)이 작업을 수행하기 위하여 취하는 일련의 프로그래밍된 행동 및 단계들이다. 디지털 어시스턴트 시스템의 능력의 범위는 작업 흐름 모델(354) 내에서 구현되

고 저장된 작업 흐름들의 수와 다양성에 의존하거나, 또는 다시 말하면, 디지털 어시스턴트 시스템(300)이 인식하는 "행동가능한 의도"의 수와 다양성에 의존한다. 그러나, 디지털 어시스턴트 시스템(300)의 실효성은 자연 언어로 표현된 사용자 요청으로부터 올바른 "행동가능한 의도(들)"를 추론하기 위한 디지털 어시스턴트 시스템의 능력에도 의존한다.

[0067] 일부 구현예들에서, 스피치-텍스트 프로세싱 모듈(330)로부터 획득된 단어들 또는 토큰들의 시퀀스뿐만 아니라, 자연 언어 프로세서(332)는 또한 (예컨대, I/O 프로세싱 모듈(328)로부터) 사용자 요청과 관련된 콘텍스트 정보를 수신한다. 자연 언어 프로세서(332)는, 스피치-텍스트 프로세싱 모듈(330)로부터 수신된 토큰 시퀀스에 포함된 정보를 명확히 하고, 보충하고/하거나 더 정의하기 위해 콘텍스트 정보를 선택적으로 사용한다. 콘텍스트 정보는, 예를 들어, 사용자 선호사항, 사용자 디바이스의 하드웨어 및/또는 소프트웨어 상태, 사용자 요청 전, 요청 중, 또는 요청 직후에 수집된 센서 정보, 디지털 어시스턴트와 사용자 간의 이전 상호작용(예를 들어, 대화) 등을 포함한다.

[0068] 일부 구현예들에서, 자연 언어 프로세싱은 온톨로지(360)에 기초한다. 온톨로지(360)는 복수의 노드들을 포함하고, 각각의 노드가 하나 이상의 "행동가능한 의도들" 또는 다른 "속성들"과 관련된 "행동가능한 의도" 또는 "속성"을 나타내는 계층적 구조이다. 전술된 바와 같이, "행동가능한 의도"는 디지털 어시스턴트 시스템(300)이 수행할 수 있는 작업(예를 들어, "행동가능"하거나 작용될 수 있는 작업)을 나타낸다. "속성"은 행동가능한 의도와 연관된 파라미터 또는 다른 속성의 하위-양태를 나타낸다. 온톨로지(360)에서 행동가능한 의도 노드와 속성 노드의 연결성은 속성 노드에 의해 표현되는 파라미터가 행동가능한 의도 노드에 의해 표현되는 작업에 관련되는 방식을 정의한다. 일부 구현예들에서, 온톨로지(360)는 행동가능한 의도 노드들 및 속성 노드들로 구성된다. 온톨로지(360) 내에서, 각각의 행동가능한 의도 노드는 하나 이상의 속성 노드들에 직접적으로 또는 하나 이상의 중간 속성 노드들을 통해 링크된다. 마찬가지로, 각각의 속성 노드는 하나 이상의 행동가능한 의도 노드들에 직접적으로 또는 하나 이상의 중간 속성 노드들을 통해 링크된다. 예를 들어, 도 3c에 도시된 온톨로지(360)는 행동가능한 의도 노드인 "레스토랑 예약" 노드를 포함한다. 속성 노드들 "레스토랑", (예약에 대한) "날짜/시간" 및 "인원수"가 각각 "레스토랑 예약" 노드(즉, 행동가능한 의도 노드)에 직접 링크된다. 또한, 속성 노드들 "요리", "가격대", "전화 번호" 및 "위치"는 속성 노드 "레스토랑"의 하위-노드들이며, 각각이 중간 속성 노드 "레스토랑"을 통해 "레스토랑 예약" 노드(즉, 행동가능한 의도 노드)에 링크된다. 다른 예에서, 도 3c에 도시된 온톨로지(360)는 또한 다른 행동가능한 의도 노드인 "리마인더 설정" 노드를 포함한다. 속성 노드들 (리마인더를 설정하기 위한) "날짜/시간" 및 (리마인더를 위한) "제목"이 각각 "리마인더 설정" 노드에 링크된다. 속성 "날짜/시간"이 레스토랑 예약 작업 및 리마인더 설정 작업 모두와 관련되기 때문에, 속성 노드 "날짜/시간"은 온톨로지(360)에서 "레스토랑 예약" 노드 및 "리마인더 설정" 노드 모두에 링크된다.

[0069] 자신이 링크된 개념 노드들과 함께, 행동가능한 의도 노드는 "도메인"으로서 기술될 수 있다. 본 논의에서, 각각의 도메인은 각각의 행동가능한 의도와 연관되며, 특정 행동가능한 의도와 연관된 노드들의 그룹(및 그들 간의 관계)을 지칭한다. 예를 들어, 도 3c에 도시된 온톨로지(360)는 온톨로지(360) 내의 레스토랑 예약 도메인(362)의 일례 및 리마인더 도메인(364)의 일례를 포함한다. 레스토랑 예약 도메인은 행동가능한 의도 노드 "레스토랑 예약", 속성 노드들 "레스토랑", "날짜/시간" 및 "인원수" 및 하위-속성 노드들 "요리", "가격대", "전화 번호" 및 "위치"를 포함한다. 리마인더 도메인(364)은 행동가능한 의도 노드 "리마인더 설정" 및 속성 노드들 "제목" 및 "날짜/시간"을 포함한다. 일부 구현예들에서, 온톨로지(360)는 다수의 도메인들로 구성된다. 각각의 도메인은 하나 이상의 속성 노드들을 하나 이상의 다른 도메인들과 공유할 수도 있다. 예를 들어, "날짜/시간" 속성 노드는 레스토랑 예약 도메인(362) 및 리마인더 도메인(364)에 더하여 다수의 다른 도메인들(예를 들어, 스케줄링 도메인, 여행 예약 도메인, 영화 티켓 도메인 등)과 연관될 수 있다. 도 3c가 온톨로지(360) 내의 두 개의 예시적인 도메인들을 도시하지만, 온톨로지(360)는 "전화 통화 개시", "방향 찾기", "미팅 스케줄링", "메시지 전송" 및 "질문에 대한 답변 제공", "사진 태그" 등과 같은 다른 도메인들(또는 행동가능한 의도들)을 포함할 수 있다. 예를 들어, "메시지 전송" 도메인은 "메시지 전송" 행동가능한 의도 노드와 연관되고, "수신자(들)", "메시지 타입" 및 "메시지 내용"과 같은 속성 노드들을 추가로 포함할 수 있다. 속성 노드 "수신자"는 예를 들어 "수신자 이름" 및 "메시지 주소"와 같은 하위-속성 노드들에 의해 추가로 정의될 수 있다.

[0070] 일부 구현예들에서, 온톨로지(360)는 디지털 어시스턴트가 이해할 수 있고 영향을 미칠 수 있는 모든 도메인들(및 이에 따른 행동가능한 의도들)을 포함한다. 일부 구현예들에서, 온톨로지(360)는, 예컨대 온톨로지(360) 내의 도메인들 또는 노드들을 추가하거나 삭제함으로써 또는 노드들 간의 관계를 수정함으로써 수정될 수 있다.

[0071] 일부 구현예들에서, 다수의 관련된 행동가능한 의도들과 연관된 노드들이 온톨로지(360) 내의 "상위 도메인" 하에서 클러스터될 수 있다. 예를 들어, "여행" 상위 도메인이 여행과 관련된 속성 노드들 및 행동가능한 의도

노드들의 클러스터를 포함할 수 있다. 여행과 관련된 행동가능한 의도 노드들은 "항공권 예약", "호텔 예약", "차량 렌트", "방향 잡기", "관심 지점 찾기" 등을 포함할 수 있다. 동일한 상위 도메인(예를 들어, "여행" 상위 도메인) 하의 행동가능한 의도 노드들은 다수의 속성 노드들을 공통으로 가질 수 있다. 예를 들어, "항공권 예약", "호텔 예약", "차량 렌트", "방향 잡기", "관심 지점 찾기"에 대한 행동가능한 의도 노드들은 하나 이상의 속성 노드들 "시작 위치", "목적지", "출발 날짜/시간", "도착 날짜/시간" 및 "인원수"를 공유할 수 있다.

[0072] 일부 구현예들에서, 온톨로지(360)에서 각각의 노드는 노드에 의해 표현되는 속성 또는 행동가능한 의도와 관련된 단어 및/또는 구들의 세트와 연관된다. 각각의 노드와 연관된 단어 및/또는 구들의 개별 세트는 노드와 연관된 이른바 "어휘"이다. 각각의 노드와 연관된 단어 및/또는 구들의 개별 세트는 노드에 의해 표현되는 속성 또는 행동가능한 의도와 연관되어 어휘 인덱스(344)(도 3b)에 저장될 수 있다. 예를 들어, 도 3b로 돌아오면, "레스토랑"의 속성에 대한 노드와 연관된 어휘가 "음식", "음료", "요리", "배고픔", "먹다", "피자", "패스트 푸드", "식사" 등과 같은 단어들을 포함할 수 있다. 다른 예를 들면, "전화 통화 개시"의 행동가능한 의도에 대한 노드와 연관된 어휘는 "통화", "전화", "다이얼", "신호음", "이 번호로 전화", "-에게 전화 걸기" 등과 같은 단어 및 구들을 포함할 수 있다. 어휘 인덱스(344)는 상이한 언어들의 단어들과 구들을 선택적으로 포함한다. 일부 구현예들에서, 도 3b에 도시된 자연 언어 프로세서(332)는 스피치-텍스트 프로세싱 모듈(330)로부터 토큰 시퀀스(예컨대, 텍스트 스트링)를 수신하고, 토큰 시퀀스 내의 단어들에 의해 어떤 노드들이 연루되는지를 결정한다. 일부 구현예들에서, 만약 토큰 시퀀스 내의 단어 또는 구가 (어휘 인덱스(344)를 통해) 온톨로지(360) 내의 하나 이상의 노드와 연관되는 것으로 발견되면, 단어 또는 구는 이러한 노드들을 "트리거" 또는 "활성화"할 것이다. 다수의 노드들이 활성화된 노드들의 수 및/또는 상대적인 중요도에 기초하여 "트리거된" 경우에, 자연 언어 프로세서(332)는 디지털 어시스턴트가 수행하도록 사용자가 의도한 작업(또는 작업 유형)으로서 행동가능한 의도들 중 하나를 선택할 것이다. 일부 구현예들에서, 가장 "트리거된" 노드들을 갖는 도메인이 선택된다. 일부 구현예들에서, (예컨대, 그의 다양한 트리거된 노드들의 상대적 중요도에 기초하여) 최고 신뢰 값을 갖는 도메인이 선택된다. 일부 구현예들에서, 도메인은 트리거된 노드들의 수 및 중요도의 조합에 기초하여 선택된다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)이 사용자로부터의 유사한 요청을 이전에 정확하게 해석했는지 여부와 같은 추가 인자들이 또한 노드를 선택하는 데 있어 고려된다.

[0073] 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 또한 어휘 인덱스(344) 내의 특정 엔티티(entity)들의 이름들을 저장하여, 이들 이름들 중 하나가 사용자 요청에서 검출되는 경우, 자연 언어 프로세서(332)는 이름이 온톨로지 내의 속성 또는 하위-속성의 특정 사례를 지칭한다는 것을 인식할 수 있을 것이다. 일부 구현예들에서, 특정 엔티티들의 이름은 사업체, 레스토랑, 사람, 영화 등의 이름이다. 일부 구현예들에서, 디지털 어시스턴트 시스템(300)은 사용자의 주소록 또는 연락처 목록, 영화 데이터베이스, 음악가 데이터베이스 및/또는 레스토랑 데이터베이스와 같은 다른 데이터 소스들로부터 특정 엔티티 이름을 검색할 수 있고 식별할 수 있다. 일부 구현예들에서, 토큰 시퀀스 내의 단어가 (사용자 주소록 또는 연락처 목록 내의 이름과 같은) 특정 엔티티의 이름임을 자연 언어 프로세서(332)가 식별하는 경우, 그 단어는 사용자 요청에 대한 온톨로지 내의 행동가능한 의도를 선택함에 있어서 추가적인 중요성을 부여받는다. 예를 들어, "Mr. Santo"라는 단어들이 사용자 요청으로부터 인식된 경우, 성인 "Santo"가 사용자의 연락처 목록 내의 연락처들 중 하나로서 어휘 인덱스(344) 내에서 발견되며, 그 다음 사용자 요청이 "메시지 전송" 또는 "전화 통화 개시" 도메인에 대응할 가능성이 있다. 다른 예를 들면, "ABC 카페"라는 단어들이 사용자 요청에서 발견된 경우, 용어 "ABC 카페"가 사용자의 도시 내의 특정 레스토랑의 이름으로서 어휘 인덱스(344)에서 발견되며, 그 다음 사용자 요청이 "레스토랑 예약" 도메인에 대응할 가능성이 있다.

[0074] 사용자 데이터(348)는, 사용자 특정 어휘, 사용자 선호사항, 사용자 주소, 사용자의 디폴트 언어와 제2 언어, 사용자의 연락처 목록 및 각각의 사용자에 대한 다른 단기 또는 장기 정보와 같은 사용자 특정 정보를 포함한다. 자연 언어 프로세서(332)는 사용자 특정 정보를 사용해서 사용자 입력에 포함된 정보를 보충하여 사용자 의도를 추가적으로 정의한다. 예를 들어, "나의 친구들을 나의 생일 파티에 초대해주세요"라는 사용자 요청에 대해서, 사용자가 자신의 요청 내에 이러한 정보를 명확하게 제공할 것을 요구하지 않고 자연 언어 프로세서(332)가 "친구들"이 누구이며 "생일 파티"가 언제 어디에서 열리는지를 결정하기 위해 사용자 데이터(348)에 액세스할 수 있다.

[0075] 일부 구현예들에서, 자연 언어 프로세서(332)는 분류 모듈(349)을 포함한다. 일부 구현예들에서, 분류 모듈(349)은, 아래에서 더 상세히 논의되는 바와 같이, (예컨대, 디지털 사진과 관련된 스피치 입력에 대응하는) 텍스트 스트링 내의 하나 이상의 용어들 각각이 엔티티, 활동성, 또는 위치 중 하나인지 여부를 결정한다. 일부 구현예들에서, 분류 모듈(349)은 하나 이상의 용어들의 각각의 용어를 엔티티, 활동성, 또는 위치 중 하나로서

분류한다. 자연 언어 프로세서(332)가 사용자 요청에 기초하여 행동가능한 의도(또는 도메인)를 식별하면, 자연 언어 프로세서(332)는 식별된 행동가능한 의도를 나타내기 위하여 구조화된 쿼리(structured query)를 생성한다. 일부 구현예들에서, 구조화된 쿼리는, 행동가능한 의도를 위한 도메인 내의 하나 이상의 노드에 대한 파라미터들을 포함하고, 파라미터들 중 적어도 일부에는 사용자 요청에 명시된 특정 정보 및 요건들이 기재된다. 예를 들어, 사용자는 "초밥집에 7시로 저녁 예약을 해주세요"라고 말할 수 있다. 이러한 경우에, 자연 언어 프로세서(332)는 사용자 입력에 기초하여 행동가능한 의도가 "레스토랑 예약"인 것으로 올바르게 식별하는 것이 가능할 수 있다. 온톨로지에 따라서, "레스토랑 예약" 도메인에 대해 구조화된 쿼리가 {요리}, {시간}, {날짜}, {인원수} 등과 같은 파라미터들을 포함할 수 있다. 사용자의 발언에 포함된 정보에 기초하여, 자연 언어 프로세서(332)는 레스토랑 예약 도메인에 대해 부분적인 구조화된 쿼리를 생성할 수 있으며, 여기에서 부분적인 구조화된 쿼리는 파라미터들 {요리= "초밥"} 및 {시간= "저녁 7시"}를 포함한다. 그러나, 이러한 예에서, 사용자의 발언은 도메인과 연관된 구조화된 쿼리를 완성하기에 불충분한 정보를 포함한다. 따라서, {인원수} 및 {날짜}와 같은 다른 필요한 파라미터들이 현재 이용가능한 정보에 기초하여 구조화된 쿼리에서 식별되지 않는다. 일부 구현예들에서, 자연 언어 프로세서(332)는 수신된 콘텍스트 정보를 구조화된 쿼리의 일부 파라미터들에 기재한다. 예를 들어, 만약 사용자가 "내 근처에 있는" 초밥 레스토랑을 요청하였다면, 자연 언어 프로세서(332)는 사용자 디바이스(104)로부터의 GPS 좌표를 이용하여 구조화된 쿼리 내에 {위치} 파라미터를 이주시킬 수 있다.

[0076] 일부 구현예들에서, 자연 언어 프로세서(332)는 (임의의 완성된 파라미터들을 포함하는) 구조화된 쿼리를 작업 흐름 프로세싱 모듈(336)("작업 흐름 프로세서")로 전달한다. 작업 흐름 프로세서(336)는: 자연 언어 프로세서(332)로부터 구조화된 쿼리를 수신하는 것, 구조화된 쿼리를 완성하는 것, 사용자의 궁극적인 요청을 "완성"하기 위해 요구되는 행동들을 수행하는 것 중 하나 이상을 수행하도록 구성된다. 일부 구현예들에서, 이들 작업들을 완수하는 데 필요한 다양한 절차들이 작업 흐름 모듈(354)에서 제공된다. 일부 구현예들에서, 작업 흐름 모듈(354)은 사용자로부터 추가 정보를 획득하기 위한 절차 및 행동가능한 의도와 연관된 행동들을 수행하기 위한 작업 흐름들을 포함한다. 위에서 설명된 바와 같이, 구조화된 쿼리를 완성시키기 위해, 작업 흐름 프로세서(336)는 추가적인 정보를 얻고/얻거나 잠재적으로 중의적인 발언들을 명확하게 하도록 사용자와 추가적인 대화를 개시할 필요가 있을 수도 있다. 이러한 상호작용이 필요한 경우, 작업 흐름 프로세서(336)는 사용자와의 대화에 관여하기 위해 대화 프로세싱 모듈(334)("대화 프로세서")를 호출한다. 일부 구현예들에서, 대화 프로세싱 모듈(334)은 어떻게(그리고/또는 언제) 사용자에게 추가 정보를 요청할 것인지 결정하고, 사용자 응답들을 수신하고 프로세싱한다. 일부 구현예들에서, 질문들은 I/O 프로세싱 모듈(328)을 통해 사용자들에게 제공되고 이들로부터 답변들이 수신된다. 예를 들어, 대화 프로세싱 모듈(334)은 청각적 및/또는 시각적 출력을 통해 사용자에게 대화 출력을 표시하고, 구두의 또는 물리적(예를 들어, 터치 제스처) 응답을 통해 사용자로부터 입력을 수신한다. 위의 예시에서 계속하여, 작업 흐름 프로세서(336)가 도메인 "레스토랑 예약"과 연관된 구조화된 쿼리에 대한 "인원수" 및 "날짜" 정보를 결정하도록 대화 프로세서(334)를 호출하면, 대화 프로세서(334)는 사용자에게 전달하기 위한 "인원수는?" 및 "날짜는?"과 같은 질문들을 생성한다. 사용자로부터 대답이 수신되면, 대화 프로세싱 모듈(334)은 구조화된 쿼리에 부족한 정보를 기재하거나 정보를 작업 흐름 프로세서(336)에 전달하여 구조화된 쿼리로부터 부족한 정보를 완성한다.

[0077] 몇몇 경우들에서, 작업 흐름 프로세서(336)는 하나 이상의 중의적인 속성들을 갖는 구조화된 쿼리를 수신할 수도 있다. 예를 들어, "메시지 전송" 도메인에 대한 구조화된 쿼리는 의도된 수신자가 "Bob"임을 나타낼 수 있고, 이때 사용자가 "Bob"이라는 이름의 다수의 연락처를 가지고 있을 수 있다. 작업 흐름 프로세서(336)는 대화 프로세서(334)가 이러한 구조화된 쿼리의 속성의 중의성을 해결할 것을 요청할 것이다. 그 결과, 대화 프로세서(334)는 사용자에게 "어떤 Bob?"이라고 물어볼 수 있고, 사용자가 선택할 수 있는 "Bob"이라는 이름의 연락처들의 목록을 디스플레이(또는 판독)할 수 있다.

[0078] 일부 구현예들에서, 대화 프로세서(334)는 중의성 해결 모듈(350)을 포함한다. 일부 구현예들에서, 중의성 해결 모듈(350)은 하나 이상의 중의적 용어들(예컨대, 디지털 사진과 관련된 스피치 입력에 대응하는 텍스트 스트림 내의 하나 이상의 중의적 용어들)의 중의성을 해결한다. 일부 구현예들에서, 중의성 해결 모듈(350)은 하나 이상의 용어들 중 제1 용어가 다수의 후보 의미들을 가짐을 식별하고, 사용자에게 제1 용어에 관한 추가 정보를 프롬프트하고, 프롬프트에 응답하여 사용자로부터 추가 정보를 수신하고, 추가 정보에 따라 제1 용어에 관련된 엔티티, 활동성, 또는 위치를 식별한다.

[0079] 일부 구현예들에서, 중의성 해결 모듈(350)은 대명사들의 중의성을 해결한다. 그러한 구현예들에서, 중의성 해결 모듈(350)은 하나 이상의 용어들 중 하나를 대명사로서 식별하고, 대명사가 지칭하는 명사를 결정한다. 일

부 구현예들에서, 중의성 해결 모듈(350)은 전자 디바이스의 사용자와 관련된 연락처 목록을 사용함으로써 대명사가 지칭하는 명사를 결정한다. 대안적으로, 또는 추가로, 중의성 해결 모듈(350)은 대명사가 이전에 태그했던 디지털 사진과 관련된 이전 스피치 입력에서 식별된 엔티티, 활동성, 또는 위치의 이름으로서 지칭하는 명사를 결정한다. 대안적으로, 또는 추가로, 중의성 해결 모듈(350)은 대명사가 이전에 태그했던 디지털 사진과 관련된 이전 스피치 입력에 기초하여 식별된 사람의 이름으로서 지칭하는 명사를 결정한다. 일부 구현예들에서, 중의성 해결 모듈(350)은 용어들 중 하나 이상의 것의 의미를 결정하기 위해 핸드헬드 전자 디바이스(예컨대, 사용자 디바이스(104))의 하나 이상의 센서들(예컨대, 근접 센서(214), 광 센서(212), GPS 수신기(213), 온도 센서(215) 및 모션 센서(210))로부터 획득된 정보에 액세스한다. 일부 구현예들에서, 중의성 해결 모듈(350)은 엔티티, 활동성, 또는 위치 중 하나와 각각 관련되는 두 개의 용어들을 식별한다. 예를 들어, 두 개의 용어들 중 제1의 것은 사람을 지칭하고, 두 개의 용어들 중 제2의 것은 위치를 지칭한다. 일부 구현예들에서, 중의성 해결 모듈(350)은 엔티티, 활동성, 또는 위치 중 하나와 각각 관련되는 세 개의 용어들을 식별한다.

[0080] 작업 흐름 프로세서(336)가 행동가능한 의도에 대한 구조화된 쿼리를 완성시키면, 작업 흐름 프로세서(336)는 행동가능한 의도와 관련된 궁극적인 작업을 수행하도록 진행한다. 따라서, 작업 흐름 프로세서(336)는 구조화된 쿼리에 포함된 특정 파라미터에 따라 작업 흐름 모델에서 단계 및 명령어를 실행한다. 예를 들어, "레스토랑 예약"의 행동가능한 의도에 대한 작업 흐름 모델은 레스토랑에 접속하여 특정 시간에 특정 인원수에 대해 실질적으로 예약할 것을 요청하는 단계들 및 명령어들을 포함할 수 있다. 예를 들어, {레스토랑 예약, 레스토랑 = ABC 카페, 날짜 = 3/12/2012, 시간 = 오후 7시, 인원수 = 5}와 같은 구조화된 쿼리를 이용하여, 작업 흐름 프로세서(336)가: (1) ABC 카페의 서버 또는 ABC 카페와 같은 다수의 레스토랑에 대한 예약을 수용하도록 구성된 레스토랑 예약 시스템에 로그인하는 단계, (2) 웹사이트 상에서 소정 양식에 날짜, 시간 및 인원수 정보를 입력하는 단계, (3) 양식을 제출하는 단계 및 (4) 사용자의 캘린더에 예약에 대한 캘린더 입력을 행하는 단계들을 수행할 수 있다. 아래에서 더 상세히 기술되는 다른 예에서, 작업 흐름 프로세서(336)는, 예컨대 사진 모듈(132)과 함께, 음성 입력에 응답하여 디지털 사진들을 태그하는 것 또는 검색하는 것과 관련되는 단계들 및 명령어들을 실행한다. 일부 구현예들에서, 작업 흐름 프로세서(336)는 사용자 입력 내에서 요청된 작업을 완수하기 위해 또는 사용자 입력 내에서 요청된 정보제공 답변을 제공하기 위해 서비스 프로세싱 모듈(338)("서비스 프로세서")의 지원을 사용한다. 예를 들어, 서비스 프로세서(338)는 전화 통화, 캘린더 입력 설정, 지도 탐색 호출, 사용자 디바이스 상에 설치된 다른 사용자 애플리케이션들 호출 또는 그와의 상호작용 및 제3자 서비스(예를 들어, 레스토랑 예약 포털, 소셜 네트워킹 웹사이트 또는 서비스, 은행 포털 등)의 호출 또는 그와의 상호작용을 위해 작업 흐름 프로세서(336)를 대신하여 행동할 수 있다. 일부 구현예들에서, 각각의 서비스에 의해 요구되는 프로토콜들 및 애플리케이션 프로그래밍 인터페이스(API)들은 서비스 모델들(356)의 각각의 서비스 모델에 의해 명시될 수 있다. 서비스 프로세서(338)는 서비스를 위한 적절한 서비스 모델에 액세스하고 서비스 모델에 따른 서비스에 의해 요구되는 프로토콜 및 API에 따라 서비스에 대한 요청을 생성한다.

[0081] 예를 들어, 레스토랑이 온라인 예약 서비스를 가능하게 한 경우, 레스토랑은 예약에 필요한 파라미터 및 온라인 예약 서비스에 필요한 파라미터들의 값을 전달하기 위한 API를 명시하는 서비스 모델을 제출할 수 있다. 작업 흐름 프로세서(336)에 의해 요청되는 경우, 서비스 프로세서(338)는 서비스 모델(356)에 저장된 웹 주소를 사용하여 온라인 예약 서비스와의 네트워크 접속을 설정하고, 온라인 예약 서비스의 API에 따른 포맷으로 예약에 필요한 파라미터(예를 들어, 시간, 날짜, 인원수)를 온라인 예약 인터페이스에 전송할 수 있다.

[0082] 일부 구현예들에서, 자연 언어 프로세서(332), 대화 프로세서(334) 및 작업 흐름 프로세서(336)는 사용자의 의도를 추론 및 정의하고, 사용자 의도를 더 명확히 하고 세밀하게 하도록 정보를 획득하고, 최종적으로 사용자의 의도를 이행하기 위해 응답을 생성(예를 들어, 사용자에게 출력을 제공하거나, 작업을 완수)하도록 총체적이고 반복적으로 사용된다.

[0083] 일부 구현예들에서, 사용자의 요청을 이행하는 데 필요한 모든 작업들이 수행된 후, 디지털 어시스턴트(326)는 확인 응답을 표현하고, I/O 프로세싱 모듈(328)을 통해 사용자에게 응답을 회신한다. 사용자 요청이 정보를 제공하는 대답을 요구하는 경우, 확인 응답은 사용자에게 요청된 정보를 표시한다. 일부 구현예들에서, 디지털 어시스턴트는 또한 사용자가 디지털 어시스턴트(326)에 의해 생성된 응답에 만족하는지 여부를 나타내도록 사용자에게 요청한다.

[0084] 이제 일부 구현예들에 따른 음성 트리거 시스템(400)의 컴포넌트들을 도시한 블록 다이어그램인 도 4에 집중한다. (음성 트리거 시스템(400)은 음성에만 제한되지 않으며, 본 명세서에 기술된 구현예들은 비-음성 사운드들에도 동일하게 적용된다.) 음성 트리거 시스템(400)은 전자 디바이스(104) 내의 다양한 컴포넌트들, 모듈들 및/또는 소프트웨어 프로그램들로 구성된다. 일부 구현예들에서, 음성 트리거 시스템(400)은 각각이 오디오 버스

(401)에 연결된 노이즈 검출기(402), 사운드-타입 검출기(404), 트리거 사운드 검출기(406) 및 스피치-기반 서비스(408) 및 오디오 서브시스템(226)을 포함한다. 일부 구현예들에서, 더 많거나 더 적은 수의 이러한 모듈들이 사용된다. 사운드 검출기(402, 404, 406)는 모듈들로 지칭될 수 있으며, 하드웨어(예로서, 회로, 메모리, 프로세서 등), 소프트웨어(예로서, 프로그램, 소프트웨어-온-칩, 펌웨어 등) 및/또는 본 명세서에 기술된 기능을 수행하기 위한 이들의 임의의 조합들을 포함할 수 있다. 일부 구현예들에서, 사운드 검출기들은 파선들에 의해 도 4에 도시된 바와 같이, (예를 들어, 통신 버스를 통해) 서로에게 통신상, 프로그램상, 물리적으로 및/또는 동작상 연결된다. (설명의 용이성을 위해서, 도 4는 각각의 사운드 검출기가 오직 인접한 사운드 검출기들에만 연결된 것으로 도시한다. 각각의 사운드 검출기는 임의의 다른 사운드 검출기들에도 연결될 수 있다는 것이 이해될 것이다.)

[0085] 일부 구현예들에서, 오디오 서브시스템(226)은 코덱(410), 오디오 디지털 신호 프로세서(DSP)(412) 및 메모리 버퍼(414)를 포함한다. 일부 구현예들에서, 오디오 서브시스템(226)은 하나 이상의 마이크로폰(230)(도 2) 및 하나 이상의 스피커(228)(도 2)에 연결된다. 오디오 서브시스템(226)은 프로세싱 및/또는 분석을 위해 (전화 및/또는 전화의 기저대역 서브시스템과 같은 다른 컴포넌트들 또는 모듈들뿐만 아니라) 사운드 검출기(402, 404, 406) 및 스피치-기반 서비스(408)에도 사운드 입력을 제공한다. 일부 구현예들에서, 오디오 서브시스템(226)은 적어도 하나의 마이크로폰(418) 및 적어도 하나의 스피커(420)를 포함하는 외부 오디오 시스템(416)에 연결된다.

[0086] 일부 구현예들에서, 스피치-기반 서비스(408)는 음성-기반 디지털 어시스턴트이며, 도 1 내지 도 3c를 참조하여 기술된 디지털 어시스턴트 시스템의 하나 이상의 컴포넌트 또는 기능에 대응한다. 일부 구현예들에서, 스피치-기반 서비스는 일부 구현예들에서 스피치-텍스트 서비스, 딕테이션 서비스 등이며, 노이즈 검출기(402)는 오디오 서브시스템(226)으로부터의 사운드 입력이 진폭 임계값과 같은 사전결정된 조건을 만족시키는지 여부를 결정하도록 오디오 채널을 모니터링한다. 오디오 채널은 하나 이상의 마이크로폰(230)(도 2)과 같은 하나 이상의 사운드 픽업 디바이스에 의해 수신된 오디오 정보의 스트림에 대응한다. 오디오 채널은 오디오 정보를 프로세싱 및/또는 전송하는 프로세싱의 상태 또는 특정 하드웨어와 무관한 오디오 정보를 지칭한다. 예를 들어, 오디오 채널은 마이크로폰(230)으로부터의 아날로그식 전기 임펄스(및/또는 이것이 전파되는 회로)뿐만 아니라, (예를 들어, 오디오 서브시스템(226) 및/또는 전자 디바이스(104)의 임의의 다른 오디오 프로세싱 시스템에 의한) 아날로그식 전기 임펄스의 프로세싱으로 발생한 디지털식으로 인코딩된 오디오 스트림을 지칭할 수 있다.

[0087] 일부 구현예들에서, 사전결정된 조건은 사운드 입력이 사전결정된 시간의 길이 동안 소정의 볼륨 위에 있는지의 여부이다. 일부 구현예들에서, 노이즈 검출기는 (사운드-타입 검출기(404), 트리거 워드 검출기(406) 및/또는 스피치-기반 서비스(408)에 의해 수행되는 것과 같은) 다른 타입의 분석에 비교하여 상대적으로 작은 컴퓨터 및 배터리 리소스를 요구하는 사운드 입력의 시간-도메인 분석을 사용한다. 일부 구현예들에서, 예를 들어 주파수-도메인 분석을 포함하는 다른 타입의 신호 프로세싱 및/또는 오디오 분석이 사용된다. 만약 노이즈 검출기(402)가 사운드 입력이 사전결정된 조건을 만족시킨다고 결정한다면, (하나 이상의 프로세싱 루틴을 개시하기 위해 제어 신호를 제공함으로써 및/또는 업스트림 사운드 검출기에 전력을 제공함으로써) 사운드-타입 검출기(404)와 같은 업스트림 사운드 검출기를 개시한다. 일부 구현예들에서, 업스트림 사운드 검출기는 다른 조건들이 만족되는 것에 응답하여 개시된다. 예를 들어, 일부 구현예들에서, 업스트림 사운드 검출기는 디바이스가 (예를 들어, 광의 임계 레벨을 검출하는 광 검출기에 기초하여) 밀폐된 공간 내에 보관되어 있지 않다는 결정에 응답하여 개시된다.

[0088] 사운드-타입 검출기(404)는 사운드 입력이 사람의 음성, 휘파람, 손뼉 등의 특정 사운드와 같은 소정의 타입의 사운드에 대응하는지 여부를 결정하도록 오디오 채널을 모니터링한다. 사운드-타입 검출기(404)가 인식하도록 구성된 사운드의 타입은 음성 트리거가 인식하도록 구성된 특정 트리거 사운드(들)에 대응할 것이다. 트리거 사운드가 구술된 단어 또는 구인 구현예들에서, 사운드-타입 검출기(404)는 "음성 활동 검출기(VAD)"를 포함한다. 일부 구현예들에서, 사운드-타입 검출기(404)는 사운드 입력의 주파수-도메인 분석을 사용한다. 예를 들어, 사운드-타입 검출기(404)는 (예를 들어, 푸리에 변환(Fourier transform)을 이용하여) 수신된 사운드 입력의 스펙트로그램(spectrogram)을 생성하고, 사운드 입력이 특정 타입 또는 카테고리의 사운드(예를 들어, 사람의 스피치)에 대응할 가능성이 있는지 여부를 결정하도록 사운드 입력의 스펙트럼 컴포넌트들을 분석한다. 따라서, 트리거 사운드가 구술된 단어 또는 구인 구현예들에서, 만약 오디오 채널이 주변 사운드(예를 들어, 교통 소음)를 들지만 사람의 스피치를 들지 않는다면, VAD는 트리거 사운드 검출기(406)를 개시하지 않을 것이다. 일부 구현예들에서, 사운드-타입 검출기(404)는 임의의 다운스트림 사운드 검출기(예를 들어, 노이즈 검출기(402))의 사전결정된 조건이 만족되는 한 활동적으로 남아있게 된다. 예를 들어, 일부 구현예들에서, 사운드-

타입 검출기(404)는 사운드 입력이 (노이즈 검출기(402)에 의해 결정된 바와 같은) 사전결정된 진폭 임계값 위에 있는 사운드를 포함하는 한 활동적으로 남아있으며, 사운드가 사전결정된 임계값 아래로 하강하는 경우 비활성화된다. 일부 구현예들에서, 한번 개시되면, 사운드-타입 검출기(402)는 타이머의 만료(예를 들어, 1, 2, 5, 또는 10초, 또는 임의의 다른 적절한 지속시간 동안), 사운드-타입 검출기(404)의 온/오프 사이클의 소정의 횟수의 만료, 또는 이벤트의 발생(예를 들어, 노이즈 검출기(402) 및/또는 사운드-타입 검출기(404)에 의해 결정된 바와 같이, 제2 임계값 아래로 사운드의 진폭이 떨어지는 것)과 같은 조건이 만족될 때까지 활동적으로 남아있게 된다.

[0089] 전술된 바와 같이, 만약 사운드-타입 검출기(404)가 사운드 입력이 사운드의 사전결정된 타입에 대응한다고 결정하면, (예를 들어, 하나 이상의 프로세싱 루틴을 개시하기 위해 제어 신호를 제공함으로써 및/또는 업스트림 사운드 검출기에 전력을 제공함으로써) 트리거 사운드 검출기(406)와 같은 업스트림 사운드 검출기를 개시한다.

[0090] 트리거 사운드 검출기(406)는 사운드 입력이 소정의 사전결정된 콘텐츠의 적어도 부분(예를 들어, 트리거 단어, 구, 또는 사운드의 적어도 부분)을 포함하는지 여부를 결정하도록 구성된다. 일부 구현예들에서, 트리거 사운드 검출기(406)는 사운드 입력의 표현("입력 표현")을 트리거 단어의 하나 이상의 기준 표현에 비교한다. 만약 입력 표현이 수용가능한 신뢰도를 가지고 하나 이상의 기준 표현 중 적어도 하나에 부합하면, 트리거 사운드 검출기(406)는 (예를 들어, 하나 이상의 프로세싱 루틴을 개시하기 위해 제어 신호를 제공함으로써 및/또는 업스트림 사운드 검출기에 전력을 제공함으로써) 스피치-기반 서비스(408)를 개시한다. 일부 구현예들에서, 입력 표현 및 하나 이상의 기준 표현은 스펙트로그램(또는 이것의 수학적 표현)이며, 이것은 신호의 스펙트럼 밀도가 시간에 따라 어떻게 달라지는지를 나타낸다. 일부 구현예들에서, 표현들은 다른 타입의 오디오 서명 또는 성문(voiceprint)이다. 일부 구현예들에서, 스피치-기반 서비스(408)를 개시하는 것은 하나 이상의 회로, 프로그램 및/또는 프로세서를 스탠바이 모드로부터 불러내는 것과 사운드-기반 서비스를 호출하는 것을 포함한다. 그 다음 사운드-기반 서비스는 더욱 포괄적인 스피치 인식, 스피치-텍스트 프로세싱 및/또는 자연 언어 프로세싱을 제공할 준비가 된다. 일부 구현예들에서, 음성-트리거 시스템(400)은 음성 인증 기능을 포함하며, 그에 따라 만약 사운드 입력이 디바이스의 소유자/사용자와 같은 특정 인물의 음성에 대응하는지를 결정할 수 있다. 예를 들어, 일부 구현예들에서, 사운드-타입 검출기(404)는 사운드 입력이 인증된 사용자에게 의해 발언되었는지를 결정하기 위한 성문 감정 기술을 사용한다. 음성 인증 및 성문 감정은 미국 특허 출원 제13/053,144호에서 더욱 자세하게 기술되며, 이것은 본 출원의 출원인에게 양도되어 그 전체가 본 명세서에 참조로서 포함된다. 일부 구현예들에서, 음성 인증은 본 명세서에 기술된 임의의 사운드 검출기들(예를 들어, 노이즈 검출기(402), 사운드-타입 검출기(404), 트리거 사운드 검출기(406) 및/또는 스피치-기반 서비스(408))에 포함된다. 일부 구현예들에서, 음성 인증은 (예를 들어, 도 4의 음성 인증 모듈(428)과 같이) 위에서 나열된 사운드 검출기들과는 별개의 모듈로서 구현되며, 노이즈 검출기(402) 뒤에, 사운드-타입 검출기(404) 뒤에, 트리거 사운드 검출기(406) 뒤에, 또는 임의의 다른 적절한 위치에 동작상 위치될 수 있다.

[0091] 일부 구현예들에서, 트리거 사운드 검출기(406)는 임의의 다운스트림 사운드 검출기(들)(예를 들어, 노이즈 검출기(402) 및/또는 사운드-타입 검출기(404))의 조건이 만족되는 한 활동적으로 남아있게 된다. 예를 들어, 일부 구현예들에서, 트리거 사운드 검출기(406)는 사운드 입력이 (노이즈 검출기(402)에 의해 검출된 것과 같이) 사전결정된 임계값 위에 있는 사운드를 포함하는 한 활동적으로 남아있게 된다. 일부 구현예들에서, 이것은 사운드 입력이 (사운드-타입 검출기(404)에 의해서 검출된 것과 같이) 소정의 타입의 사운드를 포함하는 한 활동적으로 남아있게 된다. 일부 구현예들에서, 이것은 전술된 조건들이 만족되는 한 활동적으로 남아있게 된다.

[0092] 일부 구현예들에서, 한번 개시되면, 트리거 사운드 검출기(406)는 타이머의 만료(예를 들어, 1, 2, 5, 또는 10초, 또는 임의의 다른 적절한 지속시간), 트리거 사운드 검출기(406)의 온/오프 사이클의 소정의 횟수의 만료, 또는 이벤트의 발생(예를 들어, 제2 임계값 아래로 사운드의 진폭이 떨어지는 것)과 같은 조건이 만족될 때까지 활동적으로 남아있게 된다. 일부 구현예들에서, 하나의 사운드 검출기가 다른 검출기를 개시할 때, 두 사운드 검출기들 모두가 활동적으로 남아있게 된다. 그러나, 사운드 검출기들은 다양한 시간에서 활동적 또는 비활동적일 수 있으며, 업스트림 사운드 검출기들이 활동적이기 위해서 모든 다운스트림(예로서, 더 낮은 전력 및/또는 세련도) 사운드 검출기들이 활동적이어야 할 필요는 없다(또는 그들 각각의 조건들이 만족될 필요가 없다). 예를 들어, 일부 구현예들에서, 노이즈 검출기(402) 및 사운드-타입 검출기(404)가 그들 각각의 조건들이 만족된다고 결정한 후에, 트리거 사운드 검출기(406)가 개시되고, 노이즈 검출기(402) 및 사운드-타입 검출기(404) 중 하나 또는 둘 모두는 트리거 사운드 검출기(406)가 동작하는 동안 비활성화되고/되거나 스탠바이 모드에 진입한다. 다른 구현예들에서, 노이즈 검출기(402) 및 사운드-타입 검출기(404) 모두가(또는 이들 중 하나가) 트리거 사운드 검출기(406)가 동작하는 동안 활동적 상태에 머무른다. 다양한 구현예들에서, 사운드 검출기들의

서로 다른 조합들이 서로 다른 시간에 활동적이고, 활동적 또는 비활동적인지 여부는 다른 사운드 검출기들의 상태에 의존할 수 있거나, 또는 다른 사운드 검출기들의 상태에 대해 독립적일 수 있다.

[0093] 도 4가 세 개의 분리된 사운드 검출기들을 묘사하지만, 각각이 사운드 입력의 서로 다른 양태들을 검출하도록 구성되며, 더 많거나 더 적은 사운드 검출기들이 음성 트리거의 다양한 구현에서 사용된다. 예를 들어, 일부 구현예들에서, 오직 트리거 사운드 검출기(406)만이 사용된다. 일부 구현예들에서, 트리거 사운드 검출기(406)는 노이즈 검출기(402) 또는 사운드-타입 검출기(404)와 함께 사용된다. 일부 구현예들에서, 모든 검출기들(402-406)이 사용된다. 일부 구현예들에서, 추가적인 사운드 검출기들도 포함된다.

[0094] 또한, 사운드 검출기들의 서로 다른 조합들이 서로 다른 시간에서 사용될 수 있다. 예를 들어, 사운드 검출기들의 특정 조합 및 그들이 상호작용하는 방식이 디바이스의 컨텍스트 또는 동작 상태와 같은 하나 이상의 조건에 의존할 수 있다. 구체적인 예와 같이, 만약 디바이스가 플러그인 된다면(그리고 그에 따라 오로지 배터리 전력에만 의존하지 않는다면), 트리거 사운드 검출기(406)가 활동적인 반면, 노이즈 검출기(402) 및 사운드-타입 검출기(404)는 비활동적으로 남아있게 된다. 다른 예에서, 만약 디바이스가 주머니 또는 백팩 내에 있다면, 모든 사운드 검출기들이 비활동적이다. 전술된 바와 같이 사운드 검출기들을 캐스케이드함으로써, 더 낮은 전력을 요구하는 검출기들에 의해 필요할 때에만 더 많은 전력을 요구하는 검출기들이 호출되고, 전력 효율 음성 트리거링 기능이 제공될 수 있다. 전술된 바와 같이, 추가적인 전력 효율이 듀티 사이클에 따라서 하나 이상의 사운드 검출기를 동작함으로써 획득된다. 예를 들어, 일부 구현예들에서, 시간의 적어도 부분 동안 노이즈 검출기가 오프된다고 해도, 노이즈 검출기(402)는 효율적으로 연속적인 노이즈 검출을 수행하도록 듀티 사이클에 따라 동작한다. 일부 구현예들에서, 노이즈 검출기(402)는 10ms 동안 온(on)되고 90ms 동안 오프(off)된다. 일부 구현예들에서, 노이즈 검출기(402)는 20ms 동안 온되고 500ms 동안 오프된다. 다른 온 및 오프 지속시간들도 가능하다.

[0095] 일부 구현예들에서, 만약 노이즈 검출기(402)가 자신의 "온" 간격 동안 노이즈를 검출한다면, 노이즈 검출기(402)는 사운드 입력의 추가적인 프로세싱 및/또는 분석을 위해 온으로 남아있을 것이다. 예를 들어, 노이즈 검출기(402)는 이것이 사전결정된 시간의 길이(예로서, 100ms) 동안 사전결정된 진폭 위의 사운드를 검출한다면 업스트림 사운드 검출기를 개시하도록 구성될 수 있다. 따라서, 만약 노이즈 검출기(402)가 자신의 10ms "온" 간격 동안 사전결정된 진폭보다 높은 사운드를 검출하면, 즉시 "오프" 간격으로 진입하지 않을 것이다. 대신, 노이즈 검출기(402)는 활동적으로 남아있게 되며 전체 사전결정된 지속시간(예로서, 100ms) 동안 임계값을 초과하는지 여부를 결정하기 위해 사운드 입력을 계속해서 프로세싱한다.

[0096] 일부 구현예들에서, 사운드-타입 검출기(404)는 듀티 사이클에 따라 동작한다. 일부 구현예들에서, 사운드-타입 검출기(404)는 20ms 동안 온되고 100ms 동안 오프된다. 다른 온 및 오프 지속시간들도 가능하다. 일부 구현예들에서, 사운드-타입 검출기(404)는 사운드 입력이 자신의 듀티 사이클의 "온" 간격 내에서 사전결정된 타입의 사운드에 대응하는지 여부를 결정할 수 있다. 따라서, 사운드-타입 검출기(404)는 만약 사운드-타입 검출기(404)가 자신의 "온" 간격 동안 사운드가 소정의 타입임을 결정하면, 트리거 사운드 검출기(406) (및 임의의 다른 업스트림 사운드 검출기)를 개시할 것이다. 이와 달리, 일부 구현예들에서, 만약 사운드-타입 검출기(404)가 "온" 간격 동안 사전결정된 타입에 대응할 수 있는 사운드를 검출한다면, 검출기는 즉시 "오프" 간격으로 진입하지 않을 것이다. 대신, 사운드-타입 검출기(404)는 활동적으로 남아있게 되며 계속해서 사운드 입력을 프로세싱하여 그것이 사전결정된 타입의 사운드에 대응하는지 여부를 결정한다. 일부 구현예들에서, 만약 사운드 검출기가 사전결정된 타입의 사운드가 검출되었다고 결정하면, 사운드 입력을 추가로 프로세싱하고 트리거 사운드가 검출되었는지 여부를 결정하기 위해 트리거 사운드 검출기(406)를 개시한다. 노이즈 검출기(402) 및 사운드-타입 검출기(404)와 유사하게, 일부 구현예들에서, 트리거 사운드 검출기(406)는 듀티 사이클에 따라 동작한다. 일부 구현예들에서, 트리거 사운드 검출기(406)는 50ms 동안 온되고 50ms 동안 오프된다. 다른 온 및 오프 지속시간들도 가능하다. 만약 트리거 사운드 검출기(406)가 자신의 "온" 간격 동안 트리거 사운드에 대응할 수 있는 사운드가 존재한다는 것을 검출하면, 검출기는 즉시 "오프" 간격으로 진입하지 않을 것이다. 대신, 트리거 사운드 검출기(406)는 활동적으로 남아있게 되며 계속해서 사운드 입력을 프로세싱하여 그것이 트리거 사운드를 포함하는지 여부를 결정한다. 일부 구현예들에서, 만약 이러한 사운드가 검출되면, 트리거 사운드 검출기(406)는 1, 2, 5, 또는 10초, 또는 임의의 다른 적절한 지속시간과 같은 사전결정된 지속시간 동안 오디오를 프로세싱하기 위해 활동적으로 남아있게 된다. 일부 구현예들에서, 지속시간은 검출하도록 구성된 특정한 트리거 단어 또는 사운드의 길이에 기초하여 선택된다. 예를 들어, 만약 트리거 구가 "안녕, 시리(SIRI)"라면, 트리거 단어 검출기는 사운드 입력이 해당 구를 포함하는지 여부를 결정하기 위해 약 2초 동안 동작된다.

[0097] 일부 구현예들에서, 사운드 검출기들 중 일부가 듀티 사이클에 따라 동작되는 반면, 다른 사운드 검출기들은 활

동적인 경우 계속해서 동작한다. 예를 들어, 일부 구현예들에서, 오직 제1 사운드 검출기만이 듀티 사이클에 따라 동작되며(예를 들어, 도 4의 노이즈 검출기(402)), 업스트림 사운드 검출기들은 그들이 한번 개시되면 계속해서 동작된다. 일부 다른 구현예들에서, 노이즈 검출기(402) 및 사운드-타입 검출기(404)가 듀티 사이클에 따라 동작되는 반면, 트리거 사운드 검출기(406)는 계속해서 동작된다. 특정한 사운드 검출기가 계속해서 동작되는지 또는 듀티 사이클에 따라 동작되는지 여부는 디바이스의 컨텍스트 또는 동작 상태와 같은 하나 이상의 조건에 의존한다. 일부 구현예들에서, 만약 디바이스가 플러그인 되었고 오로지 배터리 전력에만 의존하지 않는다면, 모든 사운드 검출기들이 한번 개시되면 계속해서 동작한다. 다른 구현예들에서, 만약 (예로서 센서 및/또는 마이크로폰 신호에 의해 결정된 바와 같이) 디바이스가 주머니 또는 백팩 내에 있다면 노이즈 검출기(402)(또는 임의의 사운드 검출기)가 듀티 사이클에 따라 동작하지만, 디바이스가 보관되어 있지 않은 가능성이 있는 것으로 결정되는 경우에는 계속해서 동작한다. 일부 구현예들에서, 특정한 사운드 검출기가 계속해서 동작되는지 또는 듀티 사이클에 따라 동작되는지 여부는 디바이스의 배터리 충전 레벨에 의존한다. 예를 들어, 배터리 충전이 50%보다 위에 있으면 노이즈 검출기(402)가 계속해서 동작하며, 배터리 충전이 50%보다 아래에 있으면 듀티 사이클에 따라 동작한다. 일부 구현예들에서, 음성 트리거는 노이즈, 에코 및/또는 사운드 소거 기능(집합적으로 노이즈 소거로 지칭됨)을 포함한다. 일부 구현예들에서, 노이즈 소거(noise cancellation)는 오디오 서브시스템(226)에 의해 (예를 들어, 오디오 DSP(412)에 의해) 수행된다. 노이즈 소거는 이것이 사운드 검출기들에 의해 프로세싱되기 전에 사운드 입력으로부터 원치 않는 노이즈 또는 사운드를 감소시키거나 제거한다. 일부 경우들에서, 원치 않는 노이즈는 팬(fan) 또는 키보드의 클릭 소리와 같은 사용자의 환경으로부터의 배경 노이즈이다. 일부 구현예들에서, 원치 않는 노이즈는 사전결정된 진폭 또는 주파수에 있거나 또는 그보다 위, 또는 아래에 있는 임의의 사운드이다. 예를 들어, 일부 구현예들에서, 전형적인 사람의 음성 범위(예를 들어, 3000 Hz)보다 위에 있는 사운드가 신호로부터 필터링되거나 제거된다. 일부 구현예들에서, 다수의 마이크로폰(예로서, 마이크로폰(230))이 수신된 사운드의 어떤 컴포넌트들이 감소 및/또는 제거되어야만 하는지를 결정하는 것을 돕도록 사용된다. 예를 들어, 일부 구현예들에서, 오디오 서브시스템(226)은 공간 내의 단일 포인트(예를 들어, 사용자의 입)로부터 유래한 것으로 나타나는 사운드 또는 사운드 입력의 부분들을 식별하기 위한 빔 형성 기술을 사용한다. 그 다음 오디오 서브시스템(226)은 모든 마이크로폰들에 의해 동일하게 수신되는 입력 사운드(예를 들어, 임의의 특정 방향으로부터 유래한 것으로 나타나지 않는 주변 사운드)를 사운드로부터 제거함으로써 이러한 사운드에 초점을 맞춘다.

[0098] 일부 구현예들에서, DSP(412)는 디지털 어시스턴트가 동작하는 디바이스에 의해 출력되고 있는 입력 사운드를 사운드로부터 소거 또는 제거하도록 구성된다. 예를 들어, 만약 오디오 서브시스템(226)이 음악, 라디오, 팟캐스트, 음성 출력, 또는 임의의 다른 오디오 콘텐츠를 (예로서 스피커(228)를 통해서) 출력하고 있다면, DSP(412)는 마이크로폰에 의해 들리고 사운드 입력 내에 포함된 임의의 출력된 사운드를 제거한다. 따라서, 사운드 입력은 출력된 오디오로부터 자유롭다(또는 적어도 출력된 오디오를 거의 포함하지 않는다). 따라서, 사운드 검출기들에게 제공된 사운드 입력은 더욱 깨끗할 것이며, 트리거는 보다 정확할 것이다. 노이즈 소거의 양태들은 미국 특허 제7,272,224호에 더욱 자세하게 기술되었으며, 이것은 본 출원의 출원인에게 양도되어 그 전체가 본 명세서에 참조로서 포함된다.

[0099] 일부 구현예들에서, 서로 다른 사운드 검출기들이 사운드 입력이 서로 다른 방식으로 필터링 및/또는 프로세싱될 것을 요구한다. 예를 들어, 일부 구현예들에서, 노이즈 검출기(402)는 60 내지 20,000 Hz 사이의 시간-도메인 오디오 신호를 분석하도록 구성되고, 사운드-타입 검출기는 60 내지 3,000 Hz 사이의 오디오의 주파수-도메인 분석을 수행하도록 구성된다. 따라서, 일부 구현예들에서, 오디오 DSP(412)(및/또는 디바이스(104)의 다른 오디오 DSP)가 사운드 검출기들의 개별 필요성에 따라서 수신된 오디오를 사전프로세싱한다. 일부 구현예들에서, 다른 한편으로, 사운드 검출기들은 그들의 구체적인 필요성에 따라서 오디오 서브시스템(226)으로부터의 오디오를 필터링 및/또는 사전프로세싱하도록 구성된다. 이러한 경우들에서, 오디오 DSP(412)는 여전히 사운드 검출기들에게 사운드 입력을 제공하기 이전에 노이즈 소거를 수행할 수 있다. 일부 구현예들에서, 전자 디바이스의 컨텍스트가 음성 트리거를 동작할지 여부 및 어떻게 동작할지 여부를 결정하는 것을 돕도록 사용된다. 예를 들어, 디바이스가 사용자의 주머니, 지갑, 또는 백팩에 보관되어 있을 때 사용자가 음성-기반 디지털 어시스턴트와 같은 스피치-기반 서비스를 호출할 가능성이 낮을 수 있다. 또한, 사용자가 시끄러운 락 콘서트에 있을 때 스피치-기반 서비스를 호출할 가능성이 낮을 수 있다. 일부 사용자에게 있어서, 그들이 하루 중 소정의 시간대에(예를 들어, 늦은 밤) 스피치-기반 서비스를 호출할 가능성이 낮다. 다른 한편으로, 사용자가 음성 트리거를 이용하여 스피치-기반 서비스를 호출할 가능성이 더 높은 컨텍스트 또한 존재한다. 예를 들어, 일부 사용자들은 그들이 운전 중이거나, 혼자 있거나, 일하는 중인 경우 등에 음성 트리거를 사용할 가능성이 더 높을 것이다. 다양한 기술들이 디바이스의 컨텍스트를 결정하도록 사용된다. 다양한 구현예들에서, 디바이스는 디

바이스의 콘텍스트를 결정하기 위해 GPS 수신기, 광 센서, 마이크로폰, 근접 센서, 방향 센서, 관성 센서, 카메라, 통신 회로 및/또는 안테나, 충전 및/또는 전력 회로, 스위치 포지션, 온도 센서, 나침반, 가속도계, 캘린더, 사용자 선호사항 등의 컴포넌트들 또는 정보 소스들 중 임의의 하나 이상으로부터의 정보를 사용한다. 그 다음 디바이스의 콘텍스트는 음성 트리거가 동작하는지 여부 및 어떻게 동작하는지를 조정하도록 사용될 수 있다. 예를 들어, 소정의 콘텍스트에서, 음성 트리거는 해당 콘텍스트가 유지되는 한 비활성화될 것이다(또는 상이한 모드에서 동작될 것이다). 예를 들어, 일부 구현예들에서, 전화기가 사전결정된 방향에 있을 때(예를 들어, 표면 상에서 아래를 보고 놓여있을 때), 사전결정된 시간의 주기 동안(예를 들어, 밤 10시와 오전 8시 사이), (예를 들어, 스위치 위치, 모드 설정, 또는 사용자 선호사항에 기초하여) 전화기가 "묵음" 또는 "방해 금지" 모드에 있을 때, 디바이스가 실질적으로 밀폐된 공간(예를 들어, 주머니, 가방, 지갑, 서랍, 또는 글러브 박스) 내에 있을 때, (예를 들어, 근접 센서, 음향/무선/적외선 통신에 기초하여) 디바이스가 음성 트리거 및/또는 스피치-기반 서비스를 갖는 다른 디바이스 근처에 있을 때 등의 경우에서 음성 트리거는 비활성화된다. 일부 구현예들에서, 비활성화되는 대신, 음성 트리거 시스템(400)은 저전력 모드에서 동작된다(예를 들어, 노이즈 검출기(402)를 10ms의 "온" 간격 및 5초의 "오프" 간격을 갖는 듀티 사이클에 따라 동작시킨다). 일부 구현예들에서, 오디오 채널은 음성 트리거 시스템(400)이 저전력 모드에서 동작될 때 더욱 드물게 모니터링된다. 일부 구현예들에서, 음성 트리거는 정상 모드에 있을 때보다 저전력 모드에 있을 때 상이한 사운드 검출기 또는 사운드 검출기들의 조합을 사용한다. (음성 트리거는 다수의 서로 다른 모드들 또는 동작 상태들일 수 있고, 각각은 서로 다른 전력량을 사용할 수 있으며, 서로 다른 구현예들이 자신의 특정 설계에 따라 사용할 것이다.)

[0100] 다른 한편으로, 디바이스가 일부 다른 콘텍스트에 있을 때, 음성 트리거는 콘텍스트가 유지되는 한 활성화될 것이다(또는 상이한 모드에서 동작될 것이다). 예를 들어, 일부 구현예들에서, 음성 트리거는 자신이 전원에 플러그인 되어있는 동안, 전화기가 사전결정된 방향에 있는 동안(예를 들어, 표면 상에서 위를 보고 놓여있음), 사전결정된 시간의 주기 동안(예를 들어, 오전 8시와 밤 10시 사이), (예를 들어, GPS 신호, 차량과의 블루투스 접속 또는 도킹 등에 기초하여) 디바이스가 이동 중 및/또는 차량 안에 있는 경우 등에 활동적으로 남아있게 된다. 디바이스가 차량 안에 있는 경우 라이닝을 검출하는 양태들은 미국 임시특허출원 제61/657,744호에 더욱 자세하게 기술되어 있으며, 이것은 본 출원의 출원인에게 양도되어 그 전체가 본 명세서에 참조로서 포함된다. 소정의 콘텍스트를 결정하는 방법의 몇몇 구체적인 예들이 아래에 제공된다. 다양한 실시예들에서, 서로 다른 기술들 및/또는 정보 소스들이 이러한 콘텍스트들 및 다른 콘텍스트들을 검출하도록 사용된다.

[0101] 위에서 언급된 바와 같이, 음성 트리거 시스템(400)이 활동 중인지(예를 들어, 청취 중인지) 아닌지 여부는 디바이스의 물리적 방향에 의존할 수 있다. 일부 구현예들에서, 음성 트리거는 디바이스가 표면 상에서 "위를 보게(face-up)" 배치되어 있는 경우에 (예를 들어, 디스플레이 및/또는 터치스크린 표면을 볼 수 있음) 활동적이고/이거나, "아래를 보게(face-down)" 배치되어 있는 경우에 비활동적이다. 이것은 사용자에게 설정 메뉴, 스위치, 또는 버튼의 조작을 요구하지 않고 음성 트리거를 활성화 및/또는 비활성화하는 쉬운 방식을 제공한다. 일부 구현예들에서, 디바이스는 (예를 들어, 디바이스(104)의 전면 및 후면 상의 입사광의 차이에 기초하는) 광 센서, 근접 센서, 자기 센서, 가속도계, 자이로스코프, 기울기 센서, 카메라 등을 이용하여 표면 상에서 위를 보는지 또는 아래를 보는지 여부를 검출한다. 일부 구현예들에서, 다른 동작 모드, 설정, 파라미터, 또는 선호사항이 디바이스의 방향 및/또는 위치에 의해 영향을 받는다. 일부 구현예들에서, 특정 트리거 사운드, 음성 트리거의 단어, 또는 구를 청취하는 것은 디바이스의 방향 및/또는 위치에 의존한다. 예를 들어, 일부 구현예들에서, 음성 트리거는 디바이스가 일 방향에 있을 때(예를 들어, 표면 상에서 위를 보게 놓여 있을 때) 제1 트리거 단어, 구, 또는 사운드를 청취하며, 디바이스가 다른 방향에 있을 때(예를 들어, 아래를 보게 놓여 있을 때) 다른 트리거 단어, 구, 또는 사운드를 청취한다. 일부 구현예들에서, 아래를 보는 방향에 대한 트리거 구는 위를 보는 방향에 대한 것보다 더 길고/길거나 더 복잡하다. 따라서, 사용자는 음성 트리거가 더 짧거나 단순한 트리거 단어들에 대해서 더욱 빈번할 수 있는 오류 수용을 감소시키는 동안에도 여전히 동작될 수 있도록 사용자 주위에 다른 사람들이 있거나 시끄러운 환경에 있을 때 디바이스가 아래를 보게 배치할 수 있다. 특정 예로서, 위를 보는 트리거 구는 "안녕, 시리"일 수 있는 반면, 아래를 보는 트리거 구는 "안녕, 시리, 나는 앤드류야. 나를 깨워줘"일 수 있다. 더욱 긴 트리거 구는 또한 사운드 검출기들에 대한 더 넓은 음성 샘플 및/또는 프로세싱 및/또는 분석하기 위한 음성 인증을 제공하며, 따라서 음성 트리거의 정확도를 증가시키고 오류 수용을 감소시킨다.

[0102] 일부 구현예들에서, 디바이스(104)는 자신이 차량(예를 들어, 승용차) 안에 있는지 여부를 검출한다. 음성 트리거는 디바이스 및/또는 스피치-기반 서비스를 동작시키기 위해 필요한 물리적 상호작용을 감소시키는 것을 돕기 때문에, 사용자가 차량 내에 있을 때 스피치-기반 서비스를 호출하는데 특히 유리하다. 실제로, 음성-기반 디지털 어시스턴트의 이점들 중 하나는 이것이 디바이스를 바라보거나 터치하는 것이 비실용적이거나 안전하지

않을 경우에 작업을 수행하도록 사용될 수 있다는 점이다. 따라서, 음성 트리거는 사용자가 디지털 어시스턴트를 호출하기 위해서 디바이스를 터치해야 할 필요가 없도록 디바이스가 차량 내에 있을 때 사용될 수 있다. 일부 구현예들에서, 디바이스는 예컨대 블루투스 통신(또는 다른 무선 통신)을 통한, 또는 도킹 커넥터 또는 케이블을 통해 차량에 접속되고/되거나 차량과 짝을 이루었음을 검출함으로써 자신이 차량 내에 있음을 검출한다. 일부 구현예들에서, 디바이스는 (예를 들어, GPS 수신기, 가속도계 및/또는 자이로스코프를 사용하여) 디바이스의 위치 및/또는 속도를 결정함으로써 디바이스가 차량 내에 있다고 결정한다. 만약 디바이스가 차량 내에 있을 가능성이 높다고 결정되면, 차량이 시간당 20마일보다 높은 속도로 이동하고 있으며 길을 따라 이동하고 있다고 결정되기 때문에, 예를 들어 음성 트리거가 활동적으로 남아있게 되고/되거나 고전력 또는 더욱 민감한 상태에 있게 된다.

[0103] 일부 구현예들에서, 디바이스는 디바이스가 실질적으로 밀폐된 공간 내에 있는지 여부를 결정함으로써 디바이스가 (예를 들어, 주머니, 지갑, 가방, 서랍 등에) 보관되어 있는지 여부를 검출한다. 일부 구현예들에서, 디바이스는 자신이 보관되어 있음을 결정하기 위해 광 센서(예로서, 전용 주변 광 센서 및/또는 카메라)를 사용한다. 예를 들어, 일부 구현예들에서, 광 센서가 광을 거의 검출하지 않거나 전혀 검출하지 않는다면 디바이스가 보관되어 있을 가능성이 높다. 일부 구현예들에서, 하루의 시간 및/또는 디바이스의 위치 또한 고려된다. 예를 들어, 만약 광 센서가 (예컨대, 하루 중에) 높은 광 레벨이 예상될 때에 낮은 광 레벨을 검출한다면, 디바이스는 보관되어 있을 수 있고 음성 트리거 시스템(400)이 필요하지 않을 수 있다. 따라서, 음성 트리거 시스템(400)은 저전력 또는 스탠바이 상태에 있을 것이다. 일부 구현예들에서, 디바이스의 대향하는 면들 상에 위치한 센서들에 의해 검출된 광의 차는 디바이스의 위치를 결정하도록 사용될 수 있으며, 그러므로 디바이스가 보관되어 있는지 아닌지 여부를 결정하도록 사용될 수 있다. 특히, 사용자들은 디바이스가 주머니 또는 가방 안에 보관되어 있을 때보다 디바이스가 테이블 또는 표면 상에 그대로 있을 때 음성 트리거를 활성화하고자 시도할 가능성이 높다. 그러나 디바이스가 테이블 또는 데스크와 같은 표면 상에 아래를 보도록 (또는 위를 보도록) 놓여있을 때, 디바이스의 일 표면은 해당 표면에 광이 거의 또는 전혀 도달하지 않도록 폐색되는 반면, 다른 표면은 주변 광에 노출될 것이다. 따라서, 만약 디바이스의 전면 및 후면 상의 광 센서들이 뚜렷하게 다른 광 레벨을 검출한다면, 디바이스는 자신이 보관되지 않았다고 결정한다. 다른 한편으로, 만약 대향하는 면들 상의 광 센서들이 동일하거나 유사한 광 레벨을 검출한다면, 디바이스는 자신이 실질적으로 밀폐된 공간 내에 보관되었다고 결정한다. 또한, 만약 광 센서들이 모두 낮시간 동안에 (또는 디바이스가 전화기가 밝은 환경에 있을 것으로 예상할 수 있을 때) 낮은 광 레벨을 검출한다면, 디바이스는 높은 신뢰도로 자신이 보관되어 있다고 결정한다.

[0104] 일부 구현예들에서, 다른 기술들이 디바이스가 보관되었는지 여부를 결정하기 위해 (광 센서 대신 또는 광 센서에 추가로) 사용된다. 예를 들어, 일부 구현예들에서, 디바이스는 스피커 또는 트랜듀서(예로서, 스피커(228))로부터 하나 이상의 사운드(예를 들어, 톤, 클릭, 팽 등)를 방출하며, 방출된 사운드(들)의 에코를 검출하도록 하나 이상의 마이크로폰 또는 트랜듀서(예를 들어, 마이크로폰(230))를 모니터링한다. (일부 구현예들에서, 디바이스는 사람이 듣는 범위 밖에 있는 사운드와 같은 청취 불가능한 신호를 방출한다.) 에코로부터, 디바이스는 둘러싼 환경의 특징들을 결정한다. 예를 들어, 비교적 넓은 환경(예를 들어, 방 또는 차량)이 비교적 작은, 밀폐된 환경(예를 들어, 주머니, 지갑, 가방, 서랍 등)과는 다른 사운드를 반사시킬 것이다.

[0105] 일부 구현예들에서, 만약 음성 트리거 시스템(400)이 (음성 트리거 및/또는 스피치-기반 서비스를 갖는 다른 디바이스와 같은) 다른 디바이스들 부근에 있다면, 음성 트리거 시스템(400)이 다른 디바이스들 부근에 있지 않은 경우와 상이하게 동작한다. 이것은 예를 들어, 다수의 디바이스들이 함께 가까이 있을 때, 한 사람이 트리거 단어를 발언하면 다른 둘러싼 디바이스들 또한 트리거되지 않도록 음성 트리거 시스템(400)의 민감도를 섣다운하거나 감소시키는 데에 유용할 수 있다. 일부 구현예들에서, 디바이스는 RFID, 근거리 통신, 적외선/음향 신호 등을 이용하여 다른 디바이스들에 대한 근접도를 결정한다. 앞서 언급된 바와 같이, 음성 트리거는 사용자가 운전 중인 경우와 같이 디바이스가 핸드-프리 모드에서 동작되고 있는 중에 특히 유용하다. 이러한 경우들에서, 사용자들은 종종 전화를 걸거나 텍스트 입력을 디테이션하기 위해 사용자가 자신의 얼굴 근처에 디바이스를 유지시켜야 할 필요성으로부터 자유롭게 하도록, 유선 또는 무선 헤드셋, 스피커 및/또는 마이크로폰을 갖는 손목시계, 차량의 빌트인 마이크로폰 및 스피커와 같은 외부 오디오 시스템을 사용한다. 예를 들어, 무선 헤드셋 및 차량 오디오 시스템이 블루투스 통신 또는 임의의 다른 적절한 무선 통신을 이용하여 전자 디바이스에 접속할 수 있다. 그러나, 음성 트리거에 있어서 무선 액세서리를 이용한 오픈 오디오 채널을 유지하기 위해 요구되는 전력으로 인해 무선 오디오 액세서리를 통해 수신되는 오디오를 모니터링하는 것이 비효율적일 수 있다. 특히, 무선 헤드셋은 몇 시간의 연속적인 대화 시간을 제공하기에 충분한 전하를 자신의 배터리 내에 보유할 수 있으며, 따라서 단순히 주변 오디오를 모니터링하고 가능한 트리거 사운드를 기다리도록 이용하는 대신 실질적

인 통신을 위해 헤드셋이 필요한 경우에 대해 배터리를 보존하는 데에 바람직하다. 또한, 유선 외부 헤드셋 액세서리가 단독의 온-보드 마이크로폰보다 뚜렷하게 더 많은 전력을 요구할 수 있으며, 헤드셋 마이크로폰 활동성을 유지하는 것이 디바이스의 배터리 전하를 고갈시킬 것이다. 이것은 특히 무선 또는 유선 헤드셋에 의해 수신된 주변 오디오가 전형적으로 대부분의 목음의 또는 무관한 사운드를 구성할 것이라는 것을 고려할 때 사실이다. 따라서, 일부 구현예들에서, 음성 트리거 시스템(400)은 디바이스가 외부 마이크로폰에 (유선 또는 무선으로) 연결되어 있을 때에도 디바이스 상의 마이크로폰(230)으로부터의 오디오를 모니터링한다. 그 다음, 음성 트리거가 트리거 단어를 검출하면, 디바이스는 디바이스 상의 마이크로폰(230)보다는 외부 마이크로폰을 통해 (음성-기반 디지털 어시스턴트에 대한 커맨드와 같은) 후속하는 사운드 입력을 수신하기 위해 외부 마이크로폰과의 활동적 오디오 링크를 초기화한다. 그럼에도, 소정의 조건들이 만족될 때, 활동적 통신 링크가 (유선 또는 무선)을 통해 디바이스(104)에 통신상 연결될 수 있는) 외부 오디오 시스템(416)과 디바이스 사이에서 유지될 수 있으며 그에 따라 음성 트리거 시스템(400)이 디바이스 상의 마이크로폰(230)을 대신하여 (또는 그에 더하여) 외부 오디오 시스템(416)을 통해 트리거 사운드를 청취할 수 있다. 예를 들어, 일부 구현예들에서, (예를 들어, 개별 디바이스들 상의 가속도계, 자이로스코프 등에 의해 결정된 바와 같은) 전자 디바이스 및/또는 외부 오디오 시스템(416)의 동작의 특징들이 음성 트리거 시스템(400)이 디바이스 상의 마이크로폰(230) 또는 외부 마이크로폰(418)을 이용하여 주변 사운드를 모니터링해야만 하는지 여부를 결정하도록 사용된다. 특히, 디바이스와 외부 오디오 시스템(416)의 동작 사이의 차이는 외부 오디오 시스템(416)이 실질적으로 사용되는지 여부에 대한 정보를 제공한다. 예를 들어, 만약 디바이스 및 무선 헤드셋 모두가 실질적으로 동일하게 이동 중이라면(또는 이동하지 않는다면), 헤드셋이 사용되지 않거나 착용되어 있지 않음이 결정될 수 있다. 이것은 예를 들어 두 디바이스가 서로의 부근에 있으며 유휴 상태(예를 들어, 테이블 상에 놓여 있거나 주머니, 가방, 지갑, 셔랍 등에 보관되어 있음)에 있기 때문에 발생할 수 있다. 따라서, 이러한 조건 하에서, 음성 트리거 시스템(400)은 헤드셋이 실질적으로 사용되고 있지 않을 가능성이 높기 때문에 디바이스 상의 마이크로폰을 모니터링한다. 그러나 만약 무선 헤드셋과 디바이스 사이의 동작에 차이가 존재한다면, 사용자가 헤드셋을 착용중인 것으로 결정된다. 이러한 조건은 예를 들어 디바이스를 (예로서, 표면 상에 또는 가방 안에) 내려놓은 동안 헤드셋이 머리에 착용되어 있기 때문에 발생할 수 있다(착용자가 비교적 가만히 있을 때에도 적어도 소량 이동할 가능성이 높다). 이러한 조건 하에서, 헤드셋이 착용되어 있을 가능성이 높기 때문에, 음성 트리거 시스템(400)이 활동적인 통신 링크를 유지하고 디바이스 상의 마이크로폰(230) 대신 (또는 그에 더하여) 헤드셋의 마이크로폰(418)을 모니터링한다. 그리고 이러한 기술이 디바이스 및 헤드셋의 동작의 차이에 초점을 맞추기 때문에, 두 디바이스들에 대해 공통적인 동작은 상쇄될 수 있다. 이것은 예를 들어 사용자가 이동하는 차량에서 헤드셋을 이용하고 있을 때, 디바이스(예를 들어, 셀룰러폰)가 컵 홀더, 빈 좌석, 또는 사용자의 주머니 내에서 그대로 있고 헤드셋이 사용자의 머리에 착용되어 있는 경우 유용할 수 있다. 두 디바이스들에 대한 공통적인 동작(예를 들어, 차량의 동작)이 상쇄되면, 헤드셋이 사용되고 있는 가능성이 높은지 여부(또는, 헤드셋이 착용되지 않았는지 여부)를 결정하기 위해서 (만약 존재하는 경우) 디바이스에 비교하여 헤드셋의 상대적인 동작이 결정될 수 있다. 전술된 논의에서 무선 헤드셋을 지칭하지만, 유사한 기술이 유선 헤드셋에 대해서도 적용된다.

[0106] 사람들의 음성은 매우 다르기 때문에, 특정 사용자의 음성을 인식하는 것의 정확도를 향상시키기 위해 음성 트리거를 튜닝하는 것이 필요하거나 유리할 수 있다. 또한, 예를 들어 질병, 노화 또는 호르몬 변화와 관련된 자연적인 음성 변화 등으로 인해 사람들의 음성이 시간에 따라 변화할 수 있다. 따라서, 일부 구현예들에서, 음성 트리거 시스템(400)은 특정 사용자 또는 사용자들의 그룹에 대해 자신의 음성 및/또는 사운드 인식 프로파일을 적응시킬 수 있다. 전술된 바와 같이, 사운드 검출기들(예를 들어, 사운드-타입 검출기(404) 및/또는 트리거 사운드 검출기(406))이 사운드 입력(예를 들어 사용자에게 의해 제공된 사운드 또는 발언)의 표현을 하나 이상의 기준 표현에 비교하도록 구성될 수 있다. 예를 들어, 만약 입력 표현이 기준 표현에 사전결정된 신뢰도 레벨로 부합하면, 사운드 검출기는 사운드 입력이 사전결정된 타입의 사운드에 대응한다고 결정하거나(예로서, 사운드-타입 검출기(404)) 또는 사운드 입력이 사전결정된 콘텐츠를 포함한다고 결정할 것이다(예로서, 트리거 사운드 검출기(406)). 음성 트리거 시스템(400)을 튜닝하기 위해서, 일부 구현예들에서, 디바이스는 입력 표현이 비교되는 기준 표현을 조정한다. 일부 구현예들에서, 기준 표현은 음성 등록 또는 "트레이닝" 절차의 부분으로서 조정(또는 생성)되며, 이때 사용자는 디바이스가 기준 표현을 조정(또는 생성)할 수 있도록 수차례 트리거 사운드를 출력한다. 그 결과 디바이스는 해당 사람의 실제 음성을 이용하여 기준 표현을 생성할 수 있다.

[0107] 일부 구현예들에서, 디바이스는 기준 표현을 조정하기 위해 정상 사용 조건 하에서 수신되는 트리거 사운드를 사용한다. 예를 들어, 성공적인 음성 트리거링 이벤트(예를 들어, 모든 트리거링 기준을 만족시키기 위한 사운드 입력이 발견됨) 후에, 디바이스는 기준 표현을 조정 및/또는 튜닝하기 위해 사운드 입력으로부터의 정보를

사용할 것이다. 일부 구현예들에서, 소정의 신뢰도를 가지고 트리거링 기준의 전체 또는 일부를 만족시키는 것으로 결정된 사운드 입력만이 기준 표현을 조정하도록 사용된다. 따라서, 사운드 입력이 트리거 사운드에 대응하거나 트리거 사운드를 포함하는 것에 대해 음성 트리거가 더 낮은 신뢰도를 가질 때, 해당 음성 입력이 기준 표현을 조정하기 위한 목적으로 무시될 수 있다. 다른 한편으로, 일부 구현예들에서, 더 낮은 신뢰도로 음성 트리거 시스템(400)을 만족시킨 사운드 입력은 기준 표현을 조정하도록 사용된다.

[0108]

일부 구현예들에서, 디바이스(104)는 점점 더 많은 사운드 입력이 수신되면 시간에 따라 사용자의 음성에서의 미세한 변화를 수용할 수 있도록 (이러한 기술들 또는 다른 기술들을 이용하여) 기준 표현을 반복적으로 조정한다. 예를 들어, 일부 구현예들에서, 디바이스(104)(및/또는 연관된 디바이스들 또는 서비스들)가 각각의 성공적인 트리거링 이벤트 후에 기준 표현을 조정한다. 일부 구현예들에서, 디바이스(104)는 각각의 성공적인 트리거링 이벤트와 연관된 사운드 입력을 분석하여 (예를 들어, 만약 소정의 조건들이 만족되면) 입력에 기초하여 기준 표현들이 조정되어야만 하는지 여부를 결정하며, 만약 그렇게 하는 것이 적절하다면 기준 표현만을 조정한다. 일부 구현예들에서, 디바이스(104)는 시간에 따른 기준 표현의 이동 평균을 유지한다. 일부 구현예들에서, 음성 트리거 시스템(400)은 (예를 들어, 하나 이상의 사운드 검출기에 의해 결정된 바와 같은) 하나 이상의 트리거링 기준을 만족시키지 않지만, 실질적으로 권한이 부여된 사용자에게 의해 그렇게 되도록 하는 시도일 수 있는 사운드를 검출한다. 예를 들어, 음성 트리거 시스템(400)은 "안녕, 시리"와 같은 트리거 구에 응답하도록 구성될 수 있지만, 만약 사용자의 음성이 (예로서, 질병, 노화, 액센트/억양 변화 등으로 인해) 변화되었다면, 음성 트리거 시스템(400)은 디바이스를 활성화하기 위한 사용자의 시도를 인식할 수 없다. (이것은 또한 음성 트리거 시스템(400)이 디폴트 조건으로 설정되고/되거나 사용자가 사용자의 음성에 대해 음성 트리거 시스템(400)을 맞춤화하기 위해 초기화 또는 트레이닝 절차를 수행하지 않았을 때와 같이, 음성 트리거 시스템(400)이 해당 사용자의 특정 음성에 대해 적절하게 튜닝되지 않은 경우에 발생할 수 있다.) 만약 음성 트리거 시스템(400)이 음성 트리거를 활성화하기 위한 사용자의 제1 시도에 응답하지 않는다면, 사용자는 트리거 구를 반복할 가능성이 높다. 디바이스는 이러한 반복된 사운드 입력들이 서로 유사하다는 것을 검출하고/하거나, 그들이 (음성 트리거 시스템(400)이 스피치-기반 서비스를 활성화하기에 충분히 유사하지 않음에도) 트리거 구와 유사하다고 검출한다. 만약 이러한 조건들이 만족되면, 디바이스는 사운드 입력이 음성 트리거 시스템(400)을 활성화하기 위한 유효 시도에 대응한다고 결정한다. 따라서, 일부 구현예들에서, 음성 트리거 시스템(400)은 사용자에게 의한 유사한 발언이 미래의 유효 트리거로서 수용되도록 음성 트리거 시스템(400)의 하나 이상의 양태들을 조정하기 위해 이러한 수신된 사운드 입력들을 사용한다. 일부 구현예들에서, 이들 사운드 입력은 소정의 조건들 또는 조건들의 조합만이 만족될 때에만 음성 트리거 시스템(400)을 적응시키도록 사용된다. 예를 들어, 일부 구현예들에서, 사전결정된 수의 사운드 입력들이 연속으로 수신될 때(예를 들어, 2, 3, 4, 5, 또는 임의의 다른 적절한 수), 사운드 입력들이 기준 표현과 충분히 유사할 때, 사운드 입력들이 서로 충분히 유사할 때, (예를 들어, 사운드 입력들이 사전결정된 시간의 지속시간 내에 및/또는 사전결정된 간격에서 또는 그 부근에서 수신되었을 때) 사운드 입력들이 함께 근접하고/하거나, 이들 조건 또는 다른 조건들의 임의의 조합인 경우에, 음성 트리거 시스템(400)을 적응시키도록 사용된다. 일부 경우들에서, 음성 트리거 시스템(400)은 (예를 들어, 버튼 또는 아이콘을 누름으로써) 스피치-기반 서비스를 수동 개시한 것에 이어서, 하나 이상의 트리거링 기준을 만족시키지 않는 하나 이상의 사운드 입력을 검출할 수 있다. 일부 구현예들에서, 스피치-기반 서비스가 사운드 입력이 수신된 직후에 개시되었기 때문에, 음성 트리거 시스템(400)은 사운드 입력이 실질적으로 실패한 음성 트리거링 시도에 대응된다고 결정한다. 따라서, 음성 트리거 시스템(400)은 전술된 바와 같이 사용자에게 의한 발언이 미래의 유효 트리거로서 수용될 수 있도록 음성 트리거 시스템(400)의 하나 이상의 양태들을 조정하기 위해 이러한 수신된 사운드 입력을 사용한다.

[0109]

전술된 적응 기술들이 기준 표현을 조정하는 것을 지칭하지만, 트리거 사운드 검출 기술들의 다른 양태들이 기준 표현을 조정하는 것에 더하여 또는 그 대신 동일하거나 유사한 방식으로 조정될 수 있다. 예를 들어, 일부 구현예들에서, 디바이스는 예컨대 사운드 입력의 소정의 주파수 또는 주파수의 범위에 초점을 맞추고/맞추거나 소정의 주파수 또는 주파수의 범위를 삭제하기 위해, 디바이스는 사운드 입력이 필터링되는 방식 및/또는 사운드 입력에 어떤 필터가 적용되는지를 조정한다. 일부 구현예들에서, 디바이스는 입력 표현을 기준 표현과 비교하도록 사용되는 알고리즘을 조정한다. 예를 들어, 일부 구현예들에서, 입력 표현과 기준 표현 사이의 차이를 결정하기 위해 사용되는 수학적 함수의 하나 이상의 항이 변화, 추가, 또는 제거되거나, 또는 상이한 수학적 함수로 대체된다. 일부 구현예들에서, 전술된 것과 같은 적응 기술들은 음성 트리거 시스템(400)이 제공할 수 있거나 제공하도록 구성될 수 있는 것보다 더 많은 리소스를 요구한다. 특히, 사운드 검출기들은 기준 표현 및/또는 사운드 검출 알고리즘(또는 음성 트리거 시스템(400)의 임의의 다른 적절한 양태)의 반복적인 적응을 수행하기 위해 필요한 프로세서, 데이터, 또는 메모리의 양 또는 타입을 구비하지 않을 수 있거나, 또는 그에 대한

엑세스를 갖지 않을 수 있다. 따라서, 일부 구현예들에서, 하나 이상의 전술된 적응 기술이 애플리케이션 프로세서(예를 들어, 프로세서(들)(204))와 같은 더욱 강력한 프로세서에 의해, 또는 상이한 디바이스(예를 들어, 서버 시스템(108))에 의해 수행된다. 그러나, 음성 트리거 시스템(400)은 애플리케이션 프로세서가 스탠바이 모드에 있을 때에도 동작하도록 설계된다. 따라서, 애플리케이션 프로세서가 활동적이지 않고 사운드 입력을 프로세싱할 수 없을 때 음성 트리거 시스템(400)을 적응시키도록 사용되기 위한 사운드 입력이 수신된다. 따라서, 일부 구현예들에서, 사운드 입력이 수신된 후에 추가로 프로세싱 및/또는 분석될 수 있도록 디바이스에 의해 저장된다. 일부 구현예들에서, 사운드 입력은 오디오 서브시스템(226)의 메모리 버퍼(414) 내에 저장된다. 일부 구현예들에서, (애플리케이션 프로세서가 개시될 것을 요구하지 않고 데이터가 복사되거나 이동될 수 있도록 예를 들어 DMA 엔진을 이용하는 것을 포함하는) 직접 메모리 액세스(DMA) 기술을 이용하여 사운드 입력이 시스템 메모리(예를 들어, 도 2의 메모리(250)) 내에 저장된다. 그 다음 애플리케이션 프로세서가 전술된 하나 이상의 적응 기술을 실행할 수 있도록 한번 개시되면 저장된 사운드 입력이 애플리케이션 프로세서(또는 서버 시스템(108) 또는 다른 적절한 디바이스)에 제공되거나 또는 그에 의해 액세스된다. 일부 구현예들에서,

[0110] 도 5 내지 도 7은 소정의 구현예들에 따라 음성 트리거를 동작하기 위한 방법들을 나타낸 순서도이다. 이 방법들은 선택적으로, 컴퓨터 메모리 또는 비일시적 컴퓨터 판독가능한 저장 매체(예를 들어, 클라이언트 디바이스(104)의 메모리(250), 디지털 어시스턴트 시스템(300)과 연관된 메모리(302))에 저장되고 서버 시스템(108) 및/또는 사용자 디바이스(104a)를 포함하지만 이것으로 제한되지 않는 디지털 어시스턴트 시스템의 하나 이상의 컴퓨터 시스템의 하나 이상의 프로세서에 의해 실행되는 명령어들에 의해 통제된다. 컴퓨터 판독가능 저장 매체는 자기 또는 광학 디스크 저장 디바이스, 플래시 메모리와 같은 고체 상태 저장 디바이스들, 또는 다른 비휘발성 메모리 디바이스 또는 디바이스들을 포함할 수 있다. 컴퓨터 판독가능 저장 매체 상에 저장된 컴퓨터 판독가능 명령어들은 다음 중 하나 이상을 포함할 수 있다: 소스 코드, 어셈블리 언어 코드, 객체 코드, 또는 하나 이상의 프로세서들에 의해 해석되는 다른 명령어 포맷. 다양한 구현예들에서, 각각의 방법 내의 일부 동작들이 결합될 수 있고/있거나 일부 동작들의 순서가 도면에 도시된 순서로부터 변화될 수 있다. 또한, 일부 구현예들에서, 개별 도면들에 도시된 및/또는 개별 방법들과 관련하여 논의된 동작들이 다른 방법들을 형성하도록 결합될 수 있으며, 동일한 도면에 도시된 및/또는 동일한 방법과 관련하여 논의된 동작들이 서로 다른 방법들로 분리될 수 있다. 또한, 일부 구현예들에서, 방법들 내의 하나 이상의 동작이, 예를 들어 자연 언어 프로세싱 모듈(332), 대화 흐름 프로세싱 모듈(334), 오디오 서브시스템(226), 노이즈 검출기(402), 사운드-타입 검출기(404), 트리거 사운드 검출기(406), 스피치-기반 서비스(408) 및/또는 이들의 임의의 하위모듈들을 포함하는 디지털 어시스턴트 시스템(300) 및/또는 전자 디바이스(예를 들어, 사용자 디바이스(104))의 모듈들에 의해 수행된다. 도 5는 일부 구현예들에 따라 음성 트리거 시스템(예로서, 도 4의 음성 트리거 시스템(400))을 동작하는 방법(500)을 도시한다. 일부 구현예들에서, 방법(500)이 하나 이상의 프로세서 및 하나 이상의 프로세서(예를 들어, 전자 디바이스(104))에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다. 전자 디바이스는 사운드 입력을 수신한다(502). 사운드 입력은 구술된 발언(예를 들어, 단어, 구, 또는 문장), 사람이 발생시킨 사운드(예를 들어, 휘파람 소리, 혀를 차는 소리, 손가락 튕기는 소리, 손뼉 등), 또는 임의의 다른 사운드(예를 들어, 전기적으로 발생된 처프(chirp), 기계적인 노이즈 메이커 등)에 대응할 수 있다. 일부 구현예들에서, 전자 디바이스는 (예를 들어, 도 4를 참조하여 기술된 코덱(410), 오디오 DSP(412), 버퍼(414)뿐만 아니라 마이크로폰(230, 418)을 포함하는) 오디오 서브시스템(226)을 통해 사운드 입력을 수신한다.

[0111] 일부 구현예들에서, 전자 디바이스는 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정한다(504). 일부 구현예들에서, 전자 디바이스는 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하기 위해 사운드 입력에 대한 시간-도메인 분석을 적용한다. 예를 들어, 전자 디바이스는 사운드 진폭이 사전결정된 레벨에 도달하는지 여부를 결정하도록 시간 주기에 걸쳐 사운드 입력을 분석한다. 일부 구현예들에서, 만약 사운드 입력의 진폭(예컨대, 볼륨)이 사전결정된 임계값과 만나고/만나거나 사전결정된 임계값을 초과하면 임계값이 만족된다. 일부 구현예들에서, 만약 사운드 입력이 사전결정된 시간의 길이에 대한 사전결정된 임계값을 만나고/만나거나 사전결정된 임계값을 초과하면 임계값이 만족된다. 아래에서 더욱 자세하게 논의되는 바와 같이, 일부 구현예들에서, 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하는 것(504)은 제3 사운드 검출기(예를 들어, 노이즈 검출기(402))에 의해 수행된다. (제3 사운드 검출기는 이 경우에서 다른 사운드 검출기들(예를 들어, 아래에서 논의되는 제1 및 제2 사운드 검출기들)로부터 사운드 검출기를 구별하도록 사용되며, 반드시 임의의 동작상 위치 또는 사운드 검출기들의 순서를 나타내는 것은 아니다.)

[0112] 전자 디바이스는 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정한다(506). 위에서 언급된 바와 같이, 사운드들은 사운드들의 소정의 식별가능한 특징들에 기초하여 서로 다른 "타입들"로서 분류된다.

사운드 입력이 사전결정된 타입에 대응하는지 여부를 결정하는 것은 사운드 입력이 특정 타입의 특징들을 포함하거나 나타내는지 여부를 결정하는 것을 포함한다. 일부 구현예들에서, 사전결정된 타입의 사운드는 사람의 음성이다. 이러한 구현예들에서, 사운드 입력이 사람의 음성에 대응하는지 여부를 결정하는 것은 사운드 입력이 사람의 음성의 주파수 특징을 포함하는지 여부를 결정하는 것(508)을 포함한다. 아래에서 더욱 자세하게 논의되는 바와 같이, 일부 구현예들에서, 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하는 것(506)은 제1 사운드 검출기(예를 들어, 사운드-타입 검출기(404))에 의해 수행된다. 사운드 입력이 사전결정된 타입의 사운드에 대응한다는 결정에 따라, 전자 디바이스는 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정한다(510). 일부 구현예들에서, 사전결정된 콘텐츠는 하나 이상의 사전결정된 음소에 대응한다(512). 일부 구현예들에서, 하나 이상의 사전결정된 음소는 적어도 하나의 단어를 구성한다. 일부 구현예들에서, 사전결정된 콘텐츠는 사운드(예를 들어, 휘파람, 클릭, 또는 박수)이다. 일부 구현예들에서, 아래에서 논의되는 바와 같이, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하는 것(510)은 제2 사운드 검출기(예를 들어, 트리거 사운드 검출기(406))에 의해 수행된다.

[0113] 사운드 입력이 사전결정된 콘텐츠를 포함한다는 결정에 따라, 전자 디바이스는 스피치-기반 서비스를 개시한다(514). 일부 구현예들에서, 스피치-기반 서비스는 아래에서 위에서 자세하게 기술된 바와 같은 음성-기반 디지털 어시스턴트이다. 일부 구현예들에서, 스피치-기반 서비스는 스피치 입력이 텍스트로 변환되고(예를 들어, 이메일, 텍스트 메시지, 워드 프로세싱 또는 노트 필기 애플리케이션 등의) 텍스트 입력 필드 내에 포함되고/되거나 디스플레이되는 딕테이션 서비스이다. 스피치-기반 서비스가 음성-기반 디지털 어시스턴트인 구현예에서, 음성-기반 디지털 어시스턴트가 개시되면, 사용자가 디지털 어시스턴트에게 음성 입력 및/또는 커맨드를 제공할 수 있음을 나타내는 프롬프트가 사용자에게 발행된다(예를 들어, 사운드 또는 스피치 프롬프트). 일부 구현예들에서, 음성-기반 디지털 어시스턴트를 개시하는 것은 애플리케이션 프로세서(예를 들어, 도 2의 프로세서(들)(204))를 활성화하는 것, 하나 이상의 프로그램 또는 모듈(예를 들어, 도 2의 디지털 어시스턴트 클라이언트 모듈(264))을 개시하는 것, 및/또는 원격 서버 또는 디바이스(예를 들어, 도 1의 디지털 어시스턴트 서버(106))로의 접속을 확립하는 것을 포함한다.

[0114] 일부 구현예들에서, 전자 디바이스는 사운드 입력이 특정 사용자의 음성에 대응하는지 여부를 결정한다(516). 예를 들어, 사운드 입력이 권한이 부여된 디바이스의 사용자의 음성에 대응하는지 여부를 결정하기 위해 하나 이상의 음성 인증 기술이 사운드 입력에 적용된다. 음성 인증 기술은 위에서 더욱 자세하게 기술된다. 일부 구현예들에서, 음성 인증은 사운드 검출기들 중 하나(예를 들어, 트리거 사운드 검출기(406))에 의해 수행된다. 일부 구현예들에서, 음성 인증은 (임의의 적절한 하드웨어 및/또는 소프트웨어를 포함하는) 전용 음성 인증 모듈에 의해 수행된다. 일부 구현예들에서, 사운드-기반 서비스는 사운드 입력이 사전결정된 콘텐츠를 포함하며 사운드 입력이 특정 사용자의 음성에 대응한다는 결정에 응답하여 개시된다. 따라서, 예를 들어, 사운드-기반 서비스(예를 들어, 음성-기반 디지털 어시스턴트)는 오직 권한이 주어진 사용자에게 의해 트리거 단어 또는 구가 구술되었을 때에만 개시될 것이다. 트리거 사운드의 일 사용자의 발언이 다른 사용자의 음성 트리거를 활성화하지 않을 것이기 때문에, 이것은 서비스가 권한이 없는 사용자에게 의해 호출될 수 있는 가능성을 감소시키며, 다수의 전자 디바이스들이 근거리에서 있는 경우에 특히 유용할 수 있다.

[0115] 스피치-기반 서비스가 음성-기반 디지털 어시스턴트인 일부 구현예들에서, 사운드 입력이 사전결정된 콘텐츠를 포함하지만 특정 사용자의 음성에 대응하지 않는다는 결정에 응답하여, 음성-기반 디지털 어시스턴트는 제한된 액세스 모드로 개시된다. 일부 구현예들에서, 제한된 액세스 모드는 디지털 어시스턴트가 다른 방식으로 제공할 수 있는 데이터, 서비스, 및/또는 기능 중 오직 하위세트에만 디지털 어시스턴트가 액세스하는 것을 가능하게 한다. 일부 구현예들에서, (예를 들어, 디지털 어시스턴트의 권한이 없는 사용자가 캘린더, 작업 목록, 연락처, 사진, 이메일, 텍스트 메시지 등으로부터의 데이터에 액세스할 수 없도록) 제한된 액세스 모드는 기록-전용 모드에 대응한다. 일부 구현예들에서, 제한된 액세스 모드는 스피치-기반 서비스의 샌드박스된 인스턴스(sandboxed instance)에 대응하며, 그에 따라 스피치-기반 서비스가 디바이스(104) 상의 사용자 데이터(266)(도 2)와 같은, 또는 임의의 다른 디바이스 상의 사용자의 데이터(예를 들어, 도 1의 서버 시스템(108)과 같은 원격 서버 상에 저장될 수 있는 도 3a의 사용자 데이터(348))에 기록하거나 그로부터 판독하지 않을 것이다.

[0116] 일부 구현예들에서, 사운드 입력이 사전결정된 콘텐츠를 포함하고 사운드 입력이 특정 사용자의 음성에 대응한다는 결정에 응답하여, 음성-기반 디지털 어시스턴트가 특정 사용자의 이름을 포함하는 프롬프트를 출력한다. 예를 들어, 특정 사용자가 음성 인증을 통해 식별되었을 때, 음성-기반 디지털 어시스턴트는 톤, 삐 소리, 또는 비-개인화된 음성 프롬프트와 같은 보다 일반적인 프롬프트 대신, "무엇을 도와드릴까요, 피터?"와 같은 프롬프트를 출력할 수 있다. 위에서 언급된 바와 같이, 일부 구현예들에서, 제1 사운드 검출기는 사운드 입력이 사전

결정된 타입의 사운드에 대응하는지 여부를 결정하고(단계(506)), 제2 사운드 검출기는 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정한다(단계(510)). 일부 구현예들에서, 예를 들어 제1 사운드 검출기가 제2 사운드 검출기보다 덜 프로세서-집약적인 기술을 사용하기 때문에, 제1 사운드 검출기는 동작 중에 제2 사운드 검출기보다 더 적은 전력을 소비한다. 일부 구현예들에서, 제1 사운드 검출기는 사운드-타입 검출기(404)이고, 제2 사운드 검출기는 트리거 사운드 검출기(406)이며, 둘 모두가 도 4와 관련하여 위에서 논의된다. 일부 구현예들에서, 도 4를 참조하여 위에서 기술된 바와 같이, 동작 중에 제1 및/또는 제2 사운드 검출기는 듀티 사이클에 따라 주기적으로 오디오 채널을 모니터링한다.

[0117] 일부 구현예들에서, 제1 및/또는 제2 사운드 검출기는 사운드 입력의 주파수-도메인 분석을 수행한다. 예를 들어, 이들 사운드 검출기는 주파수 스펙트럼을 생성하기 위해서 또는 사운드 입력 또는 그 일부분의 스펙트럼 밀도를 결정하기 위해서 라플라스 변환(Laplace transform), Z-변환(Z-transform), 또는 푸리에 변환(Fourier transform)을 수행한다. 일부 구현예들에서, 제1 사운드 검출기는 사운드 입력이 사람의 음성의 특징인 (또는 사람의 음성의 특징인 사운드 입력의 다른 특성, 양태, 또는 속성인) 주파수를 포함하는지 여부를 결정하도록 구성된 음성-활동 검출기이다.

[0118] 일부 구현예들에서, 제1 사운드 검출기가 사전결정된 타입의 사운드 입력을 검출할 때까지 제2 사운드 검출기는 오프되거나 활동하지 않는다. 따라서, 일부 구현예들에서, 방법(500)은 사운드 입력이 사전결정된 타입에 대응한다는 결정에 응답하여 제2 사운드 검출기를 개시하는 단계를 포함한다. (다른 구현예들에서, 제2 사운드 검출기는 다른 조건들에 응답하여 개시되거나, 또는 제1 사운드 검출기로부터의 결정과 무관하게 연속적으로 동작된다.) 일부 구현예들에서, 제2 사운드 검출기를 개시하는 것은 (예를 들어, 회로, 프로세서, 프로그램, 메모리 등을 포함하는) 하드웨어 및/또는 소프트웨어를 활성화하는 것을 포함한다. 일부 구현예들에서, 제2 사운드 검출기는 자신이 개시된 후 적어도 사전결정된 길이의 시간 동안 동작된다(예를 들어, 활동하거나 오디오 채널을 모니터링한다). 예를 들어, 제1 사운드 검출기가 사운드 입력이 사전결정된 타입에 대응한다고 결정할 때(예를 들어, 사람의 음성을 포함함), 제2 사운드 검출기가 그 사운드 입력이 또한 사전결정된 콘텐츠(예로서, 트리거 단어)를 포함하는지를 결정하도록 동작된다. 일부 구현예들에서, 사전결정된 길이의 시간은 사전결정된 콘텐츠의 지속시간에 대응한다. 따라서, 만약 사전결정된 콘텐츠가 "안녕, 시리"라는 구라면, 사전결정된 길이의 시간은 해당 구가 발언되었는지를 결정하기에 충분히 길 것이다(예를 들어, 1 또는 2초, 또는 임의의 다른 적절한 지속시간). 만약 사전결정된 콘텐츠가 "안녕 시리, 일어나서 나를 도와줘"와 같이 더 길다면, 사전결정된 시간이 더 길 것이다(예를 들어, 5초, 또는 다른 적절한 지속시간). 일부 구현예들에서, 제2 사운드 검출기는 제1 사운드 검출기가 사전결정된 타입에 대응하는 사운드를 검출하는 한 동작한다. 이러한 구현예들에서, 제1 사운드 검출기가 사운드 입력 내의 사람의 스피치를 검출한다면, 제2 사운드 검출기는 이것이 사전결정된 콘텐츠를 포함하는지를 결정하도록 사운드 입력을 프로세싱할 것이다.

[0119] 위에서 언급된 바와 같이, 일부 구현예들에서, 제3 사운드 검출기(예를 들어, 노이즈 검출기(402))는 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정한다(단계(504)). 일부 구현예들에서, 제3 사운드 검출기는 동작 중에 제1 사운드 검출기보다 더 적은 전력을 소비한다. 일부 구현예들에서, 제3 사운드 검출기는 도 4와 관련하여 위에서 논의된 바와 같이, 듀티 사이클에 따라 오디오 채널을 주기적으로 모니터링한다. 또한, 일부 구현예들에서, 제3 사운드 검출기는 사운드 입력의 시간-도메인 분석을 수행한다. 일부 구현예들에서, 시간-도메인 분석은 제2 사운드 검출기에 의해 적용되는 주파수-도메인 분석보다 덜 프로세서 집약적이기 때문에, 제3 사운드 검출기는 제1 사운드 검출기보다 더 적은 전력을 소비한다.

[0120] 제1 사운드 검출기(예를 들어, 사운드-타입 검출기(404))에 의한 결정에 응답하여 제2 사운드 검출기(예를 들어, 트리거 사운드 검출기(406))를 개시하는 것과 관련하여 위에서 논의된 것과 유사하게, 일부 구현예들에서, 제1 사운드 검출기는 제3 사운드 검출기(예를 들어, 노이즈 검출기(402))에 의한 결정에 응답하여 개시된다. 예를 들어, 일부 구현예들에서, 사운드-타입 검출기(404)는 사운드 입력이 사전결정된 조건(예를 들어, 충분한 지속시간 동안 소정의 볼륨 위에 있음)을 만족시킨다는 노이즈 검출기(402)에 의한 결정에 응답하여 개시된다. 일부 구현예들에서, 제1 사운드 검출기를 개시하는 것은 (예를 들어, 회로, 프로세서, 프로그램, 메모리 등을 포함하는) 하드웨어 및/또는 소프트웨어를 활성화하는 것을 포함한다. 다른 구현예들에서, 제1 사운드 검출기는 다른 조건들에 응답하여 개시되거나 또는 계속해서 동작된다. 일부 구현예들에서, 디바이스는 사운드 입력의 적어도 일부분을 메모리에 저장한다(518). 일부 구현예들에서, 메모리는 오디오 서브시스템(226)의 버퍼(414)(도 4)이다. 저장된 사운드 입력은 디바이스에 의한 사운드 입력의 비-실시간 프로세싱을 가능하게 한다. 예를 들어, 일부 구현예들에서, 하나 이상의 사운드 검출기가 저장된 사운드 입력을 프로세싱하기 위해 저장된 사운드 입력을 판독 및/또는 수신한다. 이것은 오디오 서브시스템(226)에 의한 사운드 입력을

수신하는 도중까지 업스트림 사운드 검출기(예를 들어, 트리거 사운드 검출기(406))가 개시되지 않는 경우에 특히 유용할 수 있다. 일부 구현예들에서, 스피치-기반 서비스가 개시되면 사운드 입력의 저장된 부분이 스피치-기반 서비스에 제공된다(520). 따라서, 스피치-기반 서비스는 사운드 입력의 해당 부분이 수신된 후까지 스피치-기반 서비스가 완전히 동작하지 않는다고 해도 사운드 입력의 저장된 부분에 대해 전사, 프로세싱, 또는 다른 방식으로 동작할 수 있다. 일부 구현예들에서, 사운드 입력의 저장된 부분이 전자 디바이스의 적응 모듈에 제공된다.

[0121] 다양한 구현예들에서, 단계(516) 내지 단계(520)가 방법(500) 내의 서로 다른 위치들에서 수행된다. 예를 들어, 일부 구현예들에서, 단계(516) 내지 단계(520) 중 하나 이상이 단계(502)와 단계(504) 사이, 단계(510)와 단계(514) 사이, 또는 임의의 다른 적절한 위치에서 수행된다.

[0122] 도 6은 일부 구현예들에 따라 음성 트리거 시스템(예를 들어, 도 4의 음성 트리거 시스템(400))을 동작하는 방법(600)을 도시한다. 일부 구현예들에서, 방법(600)은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다(예를 들어, 전자 디바이스(104)). 전자 디바이스는 자신이 사전결정된 방향에 있는지 여부를 결정한다(602). 일부 구현예들에서, 전자 디바이스는 (카메라를 포함하는) 광센서, 마이크로폰, 근접 센서, 자기 센서, 가속도계, 자이로스코프, 기울기 센서 등을 이용하여 자신의 방향을 검출한다. 예를 들어, 전자 디바이스는 앞을 보는(front-facing) 카메라의 센서 상에 입사하는 광량 또는 밝기와 뒤를 보는(rear-facing) 카메라의 센서 상에 입사하는 광량 또는 밝기를 비교함으로써 표면 상에서 아래를 보도록(face-down) 또는 위를 보도록(face-up) 놓여있는지 여부를 결정한다. 만약 앞을 보는 카메라에 의해 검출된 광량 및/또는 밝기가 뒤를 보는 카메라에 의해 검출된 것보다 뚜렷하게 더 크다면, 전자 디바이스는 자신이 위를 보고 있다고 결정할 것이다. 다른 한편으로, 만약 뒤를 보는 카메라에 의해 검출된 광량 및/또는 밝기가 앞을 보는 카메라에 의해 검출된 것보다 뚜렷하게 더 크다면, 디바이스는 자신이 아래를 보고 있다고 결정할 것이다. 전자 디바이스가 사전결정된 방향에 있다는 결정에 따라, 전자 디바이스는 음성 트리거의 사전결정된 모드를 활성화한다(604). 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 아래를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 스탠바이 모드이다(606). 예를 들어, 일부 구현예들에서, 만약 스마트폰 또는 태블릿의 스크린이 아래를 보도록 테이블 또는 데스크 상에 놓여 있다면, 음성 트리거는 음성 트리거의 의도하지 않은 활성화를 방지하기 위해서 스탠바이 모드에 놓이게 된다(예를 들어, 턴-오프 된다).

[0123] 다른 한편으로, 일부 구현예들에서, 사전결정된 방향은 실질적으로 수평이고 위를 보고 있는 디바이스의 디스플레이 스크린에 대응하며, 사전결정된 모드는 청취 모드이다(608). 따라서, 예를 들어, 만약 스마트폰 또는 태블릿의 스크린이 위를 보도록 테이블 또는 데스크 상에 놓여 있다면, 음성 트리거는 트리거를 검출하였을 때 사용자에게 응답할 수 있도록 청취 모드에 놓이게 된다.

[0124] 도 7은 일부 구현예들에 따라 음성 트리거(예를 들어, 도 4의 음성 트리거 시스템(400))를 동작하는 방법(700)을 도시한다. 일부 구현예들에서, 방법(700)은 하나 이상의 프로세서 및 하나 이상의 프로세서에 의해 실행하기 위한 명령어들을 저장하는 메모리를 포함하는 전자 디바이스에서 수행된다(예를 들어, 전자 디바이스(104)). 전자 디바이스는 제1 모드에서 음성 트리거(예를 들어, 음성 트리거 시스템(400))를 동작시킨다(702) 일부 구현예들에서, 제1 모드는 정상 청취 모드이다.

[0125] 전자 디바이스는 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐색되었음을 검출함으로써 자신이 실질적으로 밀폐된 공간 내에 있는지 여부를 결정한다(704). 일부 구현예들에서, 실질적으로 밀폐된 공간은 주머니, 지갑, 가방, 서랍, 클러브박스, 서류가방 등을 포함한다.

[0126] 전술된 바와 같이, 일부 구현예들에서, 디바이스는 스피커 또는 트랜듀서로부터 하나 이상의 사운드(예를 들어, 톤, 클릭, 팡 소리 등)를 방출하고 하나 이상의 마이크로폰 또는 트랜듀서가 방출된 사운드(들)의 에코를 검출하는지를 모니터링함으로써 마이크로폰이 폐색되었는지를 검출한다. 예를 들어, 비교적 넓은 환경(예를 들어, 방 또는 차량)이 비교적 작은, 실질적으로 밀폐된 환경(예를 들어, 지갑 또는 주머니)과 다르게 사운드를 반사시킬 것이다. 따라서, 만약 디바이스가 에코(또는 에코의 부재)에 기초하여 마이크로폰(또는 사운드가 방출된 스피커)가 폐색되었음을 검출하면, 디바이스는 자신이 실질적으로 밀폐된 공간 내에 있다고 결정한다. 일부 구현예들에서, 디바이스는 마이크로폰이 밀폐된 공간의 사운드 특징을 픽업하였음을 검출함으로써 마이크로폰이 폐색되었음을 검출한다. 예를 들어, 디바이스가 주머니 내에 있을 때, 마이크로폰은 마이크로폰이 주머니의 섬유와 접촉하거나 근접하게 됨으로 인해 특징적인 바스락거리는 노이즈를 검출할 수 있다. 일부 구현예들에서, 디바이스는 센서에 의해, 또는 초점이 맞춰진 이미지를 획득할 수 있는지 여부를 결정함으로써 수신된 광의 레

벨에 기초하여 카메라가 폐색되었는지를 검출한다. 예를 들어, 만약 카메라 센서가 높은 레벨의 광이 예상될 수 있는 시간 동안 (예를 들어, 낮시간 동안) 낮은 레벨의 광을 검출한다면, 디바이스는 카메라가 폐색되었고, 그 디바이스는 실질적으로 밀폐된 공간 내에 있다고 결정한다. 다른 예로서, 카메라는 자신의 센서 상에 초점이 맞은 이미지를 획득하고자 시도할 수 있다. 일반적으로, 만약 카메라가 극도로 어두운 장소(예를 들어, 주머니 또는 백팩) 내에 있거나, 또는 카메라가 초점을 맞추고자 시도하고 있는 물체에 너무 가까이 있다면(예를 들어, 지갑 또는 백팩 내부), 초점이 맞은 이미지를 획득하는 것은 어려울 것이다. 따라서, 만약 카메라가 초점이 맞은 이미지를 획득할 수 없다면, 디바이스가 실질적으로 밀폐된 공간 내에 있다고 결정한다.

[0127] 전자 디바이스가 실질적으로 밀폐된 공간 내에 있다는 결정에 따라, 전자 디바이스는 음성 트리거를 제2 모드로 전환한다(706). 일부 구현예들에서, 제2 모드는 스탠바이 모드이다(708). 일부 구현예들에서, 스탠바이 모드에 있을 때, 음성 트리거 시스템(400)은 계속해서 주변 오디오를 모니터링할 것이지만, 다른 방식으로 음성 트리거 시스템(400)을 트리거할 것인지 여부와 무관하게 수신된 사운드에 응답하지 않을 것이다. 일부 구현예들에서, 스탠바이 모드에서는 음성 트리거 시스템(400)이 비활성화되며, 트리거 사운드를 검출하기 위해 오디오를 프로세싱하지 않는다. 일부 구현예들에서, 제2 모드는 제1 모드와 다른 듀티 사이클에 따라 음성 트리거 시스템(400)의 하나 이상의 사운드 검출기를 동작시키는 것을 포함한다. 일부 구현예들에서, 제2 모드는 제1 모드와 다른 사운드 검출기들의 조합을 동작시키는 것을 포함한다.

[0128] 일부 구현예들에서, 실질적으로 밀폐된 공간 내에 있다고 해도 음성 트리거 시스템(400)이 트리거 사운드를 검출하고 그에 반응할 수 있도록 제2 모드는 더욱 민감한 모니터링 모드에 대응한다. 일부 구현예들에서, 음성 트리거가 제2 모드로 전환되면, 디바이스는 (예를 들어, 단계(704)와 관련하여 전술된 임의의 기술을 이용하여) 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐색되었는지 여부를 검출함으로써 전자 디바이스가 여전히 실질적으로 밀폐된 공간 내에 있는지 여부를 주기적으로 결정한다. 만약 디바이스가 실질적으로 밀폐된 공간 내에 남아있다면, 음성 트리거 시스템(400)은 제2 모드로 유지될 것이다. 일부 구현예들에서, 만약 디바이스가 실질적으로 밀폐된 공간으로부터 제거되면, 전자 디바이스는 음성 트리거를 제1 모드로 복귀시킬 것이다.

[0129] 일부 구현예들에 따라, 도 8은 전술된 바와 같은 본 발명의 원리에 따라 구성된 전자 디바이스(800)의 기능 블록 다이어그램을 도시한다. 디바이스의 기능 블록들은 본 발명의 원리들을 실행하도록 하드웨어, 소프트웨어 또는 하드웨어와 소프트웨어의 조합에 의해 구현될 수 있다. 도 8에 묘사된 기능 블록들은 앞서 기재된 바와 같이 본 발명의 원리들을 구현하기 위해 조합되거나 서브-블록들로 분리될 수 있음이 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 이해된다. 따라서, 본 명세서에서의 설명은 본 명세서에서 기술된 기능 블록들의 임의의 가능한 조합 또는 분리 또는 추가적인 정의를 지원할 수 있다.

[0130] 도 8에 도시된 바와 같이, 전자 디바이스(800)는 사운드 입력을 수신하도록 구성된 사운드 수신 유닛(802)을 포함한다. 전자 디바이스(800)는 또한 스피치 수신 유닛(802)에 연결된 프로세싱 유닛(806)을 포함한다. 일부 구현예들에서, 프로세싱 유닛(806)은 노이즈 검출 유닛(808), 사운드 타입 검출 유닛(810), 트리거 사운드 검출 유닛(812), 서비스 개시 유닛(814) 및 음성 인증 유닛(816)을 포함한다. 일부 구현예들에서, 노이즈 검출 유닛(808)은 위에서 논의된 노이즈 검출기(402)에 대응하며, 노이즈 검출기(402)를 참조로 하여 전술된 임의의 동작들을 수행하도록 구성된다. 일부 구현예들에서, 사운드 타입 검출 유닛(810)은 위에서 논의된 사운드-타입 검출기(404)에 대응하며, 사운드-타입 검출기(404)를 참조로 하여 전술된 임의의 동작들을 수행하도록 구성된다. 일부 구현예들에서, 트리거 사운드 검출 유닛(812)은 위에서 논의된 트리거 사운드 검출기(406)에 대응하며, 트리거 사운드 검출기(406)를 참조로 하여 전술된 임의의 동작들을 수행하도록 구성된다. 일부 구현예들에서, 음성 인증 유닛(816)은 위에서 논의된 음성 인증 모듈(428)에 대응하며, 음성 인증 모듈(428)을 참조로 하여 전술된 임의의 동작들을 수행하도록 구성된다. 프로세싱 유닛(806)은 사운드 입력의 적어도 일부분이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하고(예를 들어, 사운드 타입 검출 유닛(810)을 이용); 사운드 입력의 적어도 일부분이 사전결정된 타입에 대응한다는 결정에 따라, 사운드 입력이 사전결정된 콘텐츠를 포함하는지 여부를 결정하며(예를 들어, 트리거 사운드 검출 유닛(812)을 이용); 사운드 입력이 사전결정된 콘텐츠를 포함한다는 결정에 따라, 스피치-기반 서비스를 개시한다(예를 들어, 서비스 개시 유닛(814)을 이용).

[0131] 일부 구현예들에서, 프로세싱 유닛(806)은 또한, 사운드 입력이 사전결정된 타입의 사운드에 대응하는지 여부를 결정하기 전에, 사운드 입력이 사전결정된 조건을 만족시키는지 여부를 결정하도록 구성된다(예를 들어, 노이즈 검출 유닛(808)을 이용). 일부 구현예들에서, 프로세싱 유닛(806)은 또한 사운드 입력이 특정 사용자의 음성에 대응하는지 여부를 결정하도록 구성된다(예를 들어, 음성 인증 유닛(816)을 이용).

[0132] 일부 구현예들에 따라서, 도 9는 전술된 바와 같은 본 발명의 원리들에 따라 구성된 전자 디바이스(900)의 기능

블록 다이어그램을 도시한다. 디바이스의 기능 블록들은 본 발명의 원리들을 실행하도록 하드웨어, 소프트웨어 또는 하드웨어와 소프트웨어의 조합에 의해 구현될 수 있다. 도 9에 묘사된 기능 블록들은 앞서 기재된 바와 같이 본 발명의 원리들을 구현하기 위해 조합되거나 서브-블록들로 분리될 수 있음이 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 이해된다. 따라서, 본 명세서에서의 설명은 본 명세서에서 기술된 기능 블록들의 임의의 가능한 조합 또는 분리 또는 추가적인 정의를 지원할 수 있다.

[0133] 도 9에 도시된 바와 같이, 전자 디바이스(900)는 음성 트리거 유닛(902)을 포함한다. 음성 트리거 유닛(902)은 다양한 서로 다른 모드들에서 동작될 수 있다. 제1 모드에서, 음성 트리거 유닛은 사운드 입력을 수신하고 사운드 입력이 소정의 기준을 만족시키는지를 결정한다(예를 들어, 청취 모드). 제2 모드에서, 음성 트리거 유닛(902)은 사운드 입력을 수신하지 않고/않거나 프로세싱하지 않는다(예를 들어, 스탠바이 모드). 전자 디바이스(900)는 또한 음성 트리거 유닛(902)에 연결된 프로세싱 유닛(906)을 포함한다. 일부 구현예들에서, 프로세싱 유닛(906)은 (예를 들어, 마이크로폰, 카메라, 가속도계, 자이로스코프 등을 포함하는) 하나 이상의 센서를 포함할 수 있고/할 수 있거나 하나 이상의 센서와 인터페이스할 수 있는 환경 검출 유닛(908) 및 모드 전환 유닛(910)을 포함한다. 일부 구현예들에서, 프로세싱 유닛(906)은: 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐쇄되었음을 검출함으로써 전자 디바이스가 실질적으로 밀폐된 공간 내에 있는지 여부를 결정하고(예를 들어, 환경 검출 유닛(908)을 이용); 전자 디바이스가 실질적으로 밀폐된 공간 내에 있다는 결정에 따라, 음성 트리거를 제2 모드로 전환하도록 구성된다(예를 들어, 모드 전환 유닛(910)을 이용).

[0134] 일부 구현예들에서, 프로세싱 유닛은: 전자 디바이스가 사전결정된 방향에 있는지 여부를 결정하고(예를 들어, 환경 검출 유닛(908)을 이용); 전자 디바이스가 사전결정된 방향에 있다는 결정에 따라, 음성 트리거의 사전결정된 모드를 활성화하도록 구성된다(예를 들어, 모드 전환 유닛(910)을 이용).

[0135] 일부 구현예들에 따라서, 도 10은 전술된 바와 같은 본 발명의 원리들에 따라 구성된 전자 디바이스(1000)의 기능 블록 다이어그램을 도시한다. 디바이스의 기능 블록들은 본 발명의 원리들을 실행하도록 하드웨어, 소프트웨어 또는 하드웨어와 소프트웨어의 조합에 의해 구현될 수 있다. 도 10에 묘사된 기능 블록들은 전술된 바와 같이 본 발명의 원리들을 구현하기 위해 조합되거나 서브-블록들로 분리될 수 있음이 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 이해된다. 따라서, 본 명세서에서의 설명은 본 명세서에서 기술된 기능 블록들의 임의의 가능한 조합 또는 분리 또는 추가적인 정의를 지원할 수 있다.

[0136] 도 10에 도시된 바와 같이, 전자 디바이스(1000)는 음성 트리거 유닛(1002)을 포함한다. 음성 트리거 유닛(1002)은 다양한 서로 다른 모드들에서 동작될 수 있다. 제1 모드에서, 음성 트리거 유닛은 사운드 입력을 수신하고 사운드 입력이 소정의 기준을 만족시키는지를 결정한다(예를 들어, 청취 모드). 제2 모드에서, 음성 트리거 유닛(1002)은 사운드 입력을 수신하지 않고/않거나 프로세싱하지 않는다(예를 들어, 스탠바이 모드). 전자 디바이스(1000)는 또한 음성 트리거 유닛(1002)에 연결된 프로세싱 유닛(1006)을 포함한다. 일부 구현예들에서, 프로세싱 유닛(1006)은 마이크로폰 및/또는 카메라를 포함할 수 있고/할 수 있거나 이들과 인터페이스할 수 있는 환경 검출 유닛(1008) 및 모드 전환 유닛(1010)을 포함한다.

[0137] 프로세싱 유닛(1006)은: 전자 디바이스의 하나 이상의 마이크로폰 및 카메라가 폐쇄되었음을 검출함으로써 전자 디바이스가 실질적으로 밀폐된 공간 내에 있는지 여부를 결정하고(예를 들어, 환경 검출 유닛(1008)을 이용); 전자 디바이스가 실질적으로 밀폐된 공간 내에 있다는 결정에 따라, 음성 트리거를 제2 모드로 전환하도록 구성된다(예를 들어, 모드 전환 유닛(1010)을 이용). 전술된 내용은 설명을 위해 특정 구현예들을 참조하여 기술되었다. 그러나, 상기의 예시적인 논의들은 개시된 구현예들을 망라하거나 또는 개시된 정확한 형태들로 제한하려는 의도는 아니다. 많은 수정들 및 변형들이 상기 교시 내용들의 관점에서 가능하다. 구현예들은 개시된 아이디어들의 원리들 및 실제 응용들을 가장 잘 설명하여, 이로써 당업자가 이들을 고려되는 특정 용도에 맞춰진 바와 같은 다양한 수정들과 함께 가장 잘 활용할 수 있게 하기 위해 선택되었고 기술되었다.

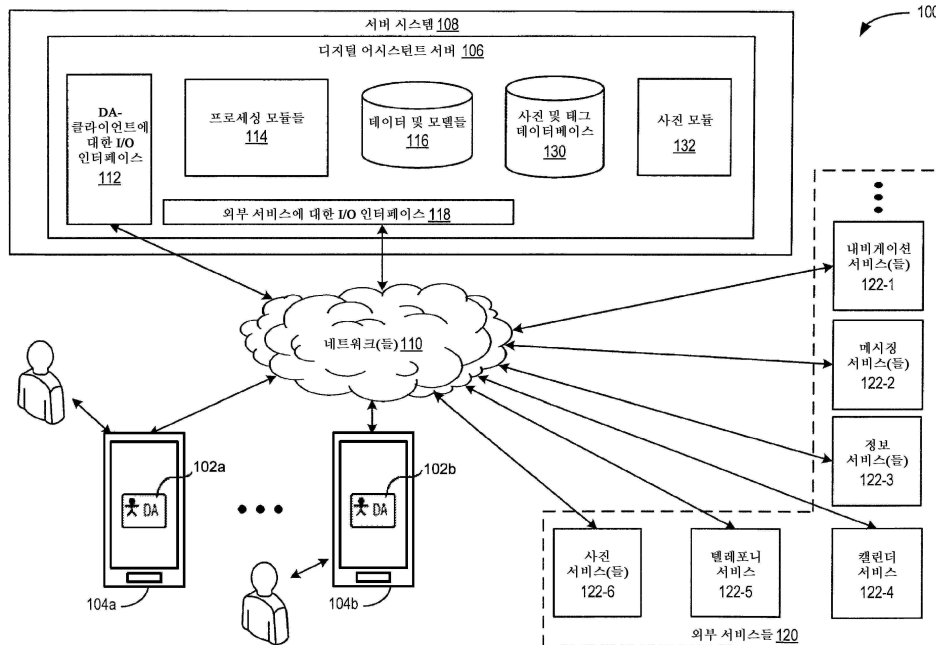
[0138] "제1", "제2" 등의 용어들이 본 명세서에서 다양한 요소들을 기술하기 위해 사용될 수 있지만, 이러한 요소들이 이들 용어들로 제한되어서는 안된다는 것이 이해될 것이다. 이들 용어들은 하나의 요소를 다른 요소와 구별하는 데에만 사용된다. 예를 들어, 등장하는 모든 제1 사운드 검출기가 일관적으로 명칭이 변경되고 등장하는 모든 제2 사운드 검출기가 일관적으로 명칭이 변경되는 한, 설명의 의미를 변화시키지 않고 제1 사운드 검출기가 제2 사운드 검출기로 명명될 수 있으며, 유사하게, 제2 사운드 검출기가 제1 사운드 검출기로 명명될 수 있다. 제1 사운드 검출기 및 제2 사운드 검출기는 모두 사운드 검출기들이지만, 동일한 사운드 검출기는 아니다.

[0139] 본 명세서에서 사용되는 용어는 단지 특정 구현예들만을 기술하기 위한 것이고, 특허청구범위를 제한하고자 의도되는 것은 아니다. 구현예들 및 첨부된 특허청구범위의 설명에서 사용되는 바와 같은 단수 형태의 명사들은

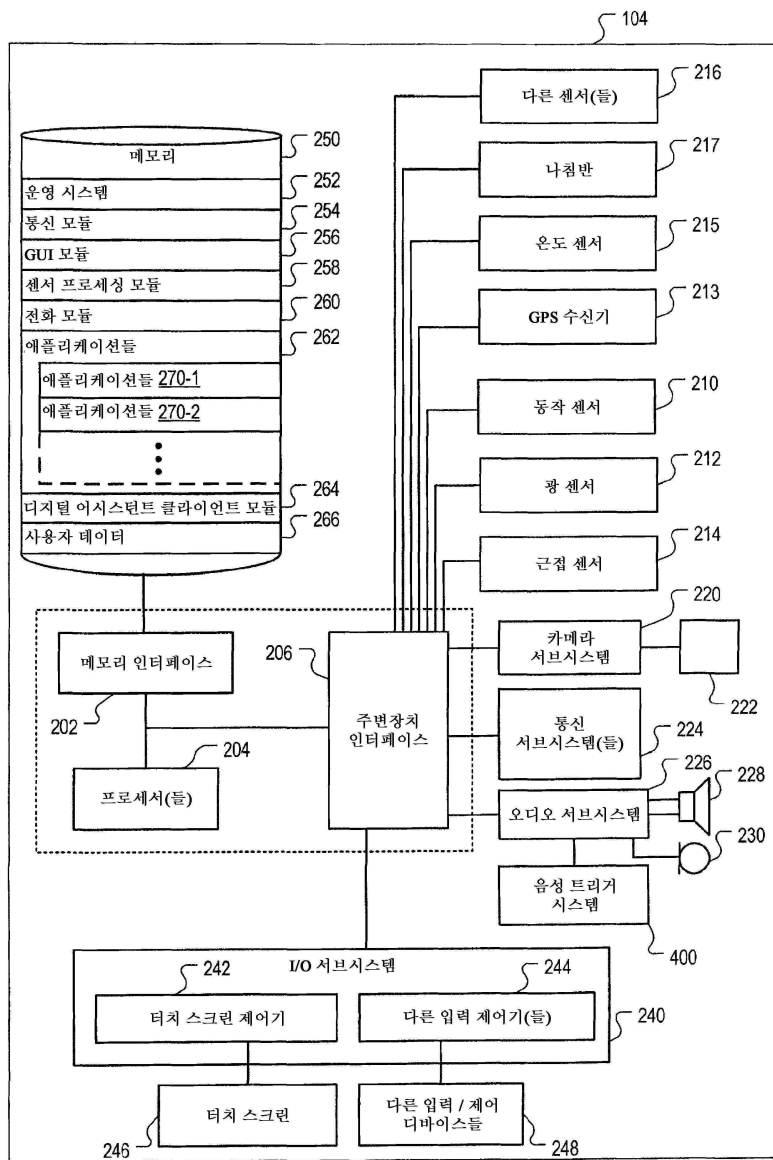
맥락에서 달리 명백하게 나타내지 않는 한, 복수 형태의 명사들 또한 포함하고자 의도된다. 또한 본 명세서에서 사용되는 바와 같은 용어 "및/또는"이 하나 이상의 연관된 나열된 아이템들의 임의의 가능한 조합 및 모든 가능한 조합들을 지칭하며 이를 망라한다는 것이 이해될 것이다. 또한 용어 "포함한다" 및/또는 "포함하는"이 본 명세서에 사용되었을 때, 이러한 용어가 언급된 특성, 정수, 단계, 동작, 요소, 및/또는 컴포넌트들의 존재를 명시하지만, 하나 이상의 다른 특성, 정수, 단계, 동작, 요소, 및/또는 컴포넌트들, 및/또는 이들의 그룹의 존재 또는 추가를 배제하는 것은 아니라는 것이 추가로 이해될 것이다. 본 명세서에서 사용되는 바와 같이, 용어 "만약 ~한다면"은 맥락에 따라, 언급된 조건 선험가 참인 "~하는 경우" 또는 "~함에 따라" 또는 "결정에 응답하여" 또는 "결정에 따라서" 또는 "검출에 응답하여"를 의미하는 것으로 해석될 수 있다. 유사하게, "만약 [언급된 조건 선험가 참이라고] 결정된다면" 또는 "만약 [언급된 조건 선험가 참]이라면" 또는 "[언급된 조건 선험가 참]인 경우"와 같은 구가 맥락에 따라 언급된 조건 선험가 참인 "결정함에 따라" 또는 "~라는 결정에 따라" 또는 "결정한 것에 응답하여" 또는 "결정에 따라서" 또는 "검출함에 따라" 또는 "검출한 것에 응답하여"를 의미하는 것으로 해석될 수 있다.

도면

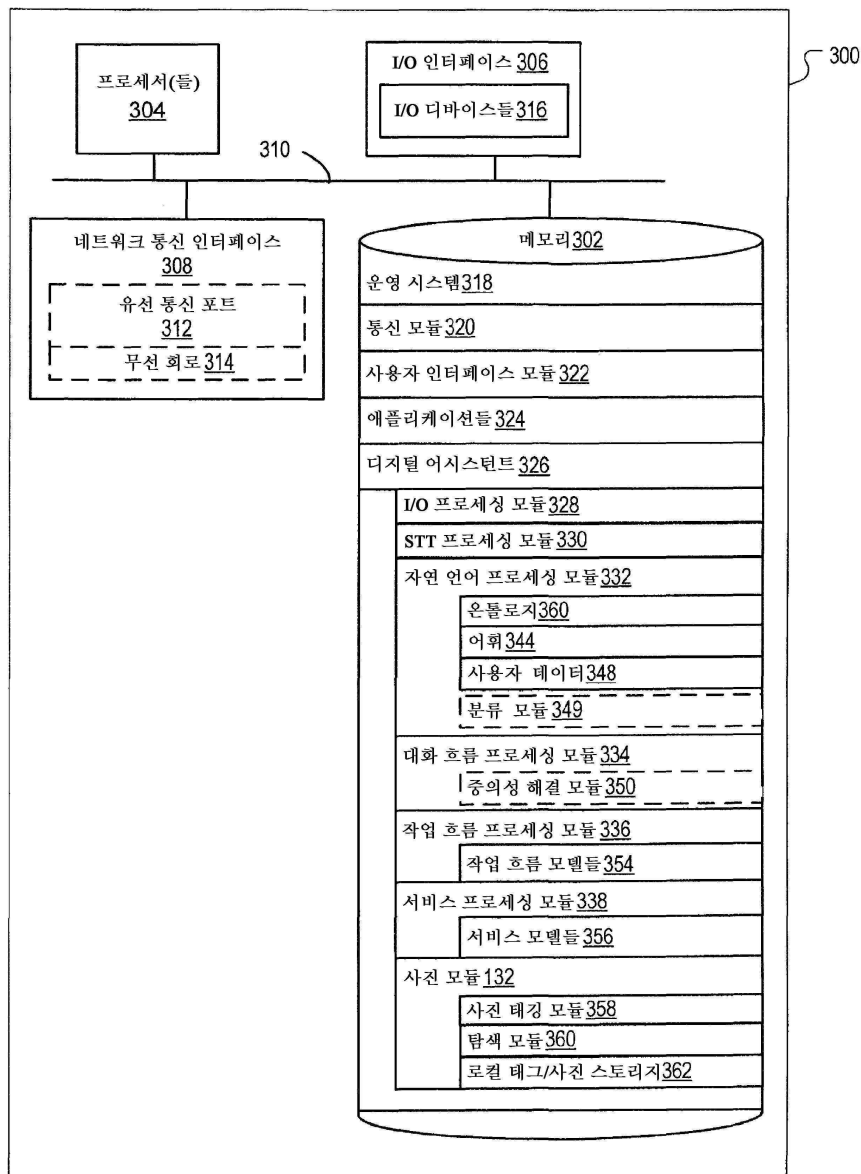
도면1



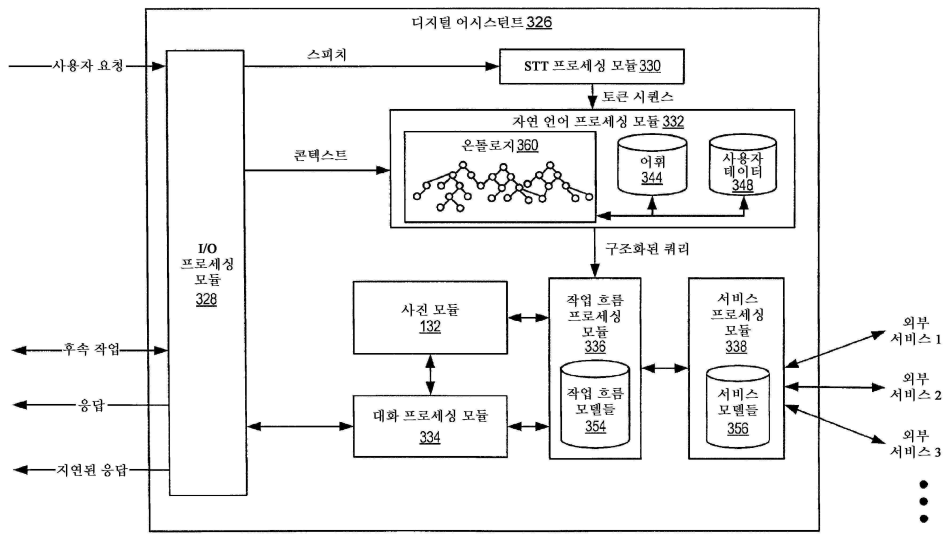
도면2



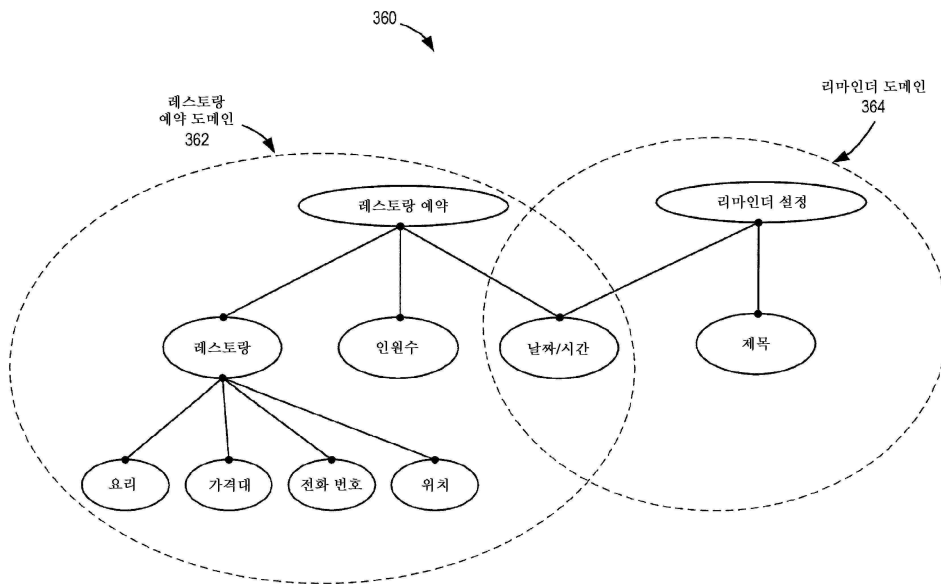
도면3a



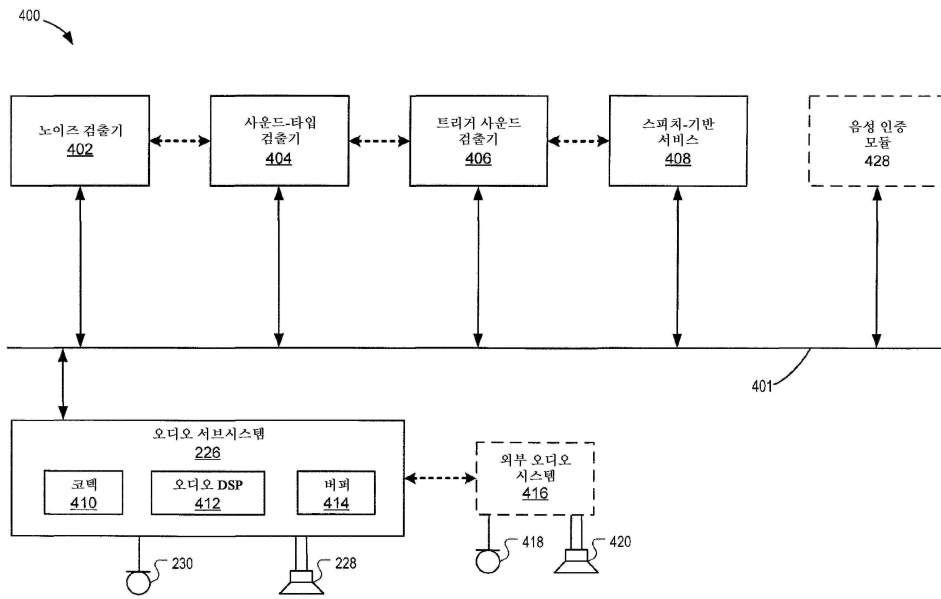
도면3b



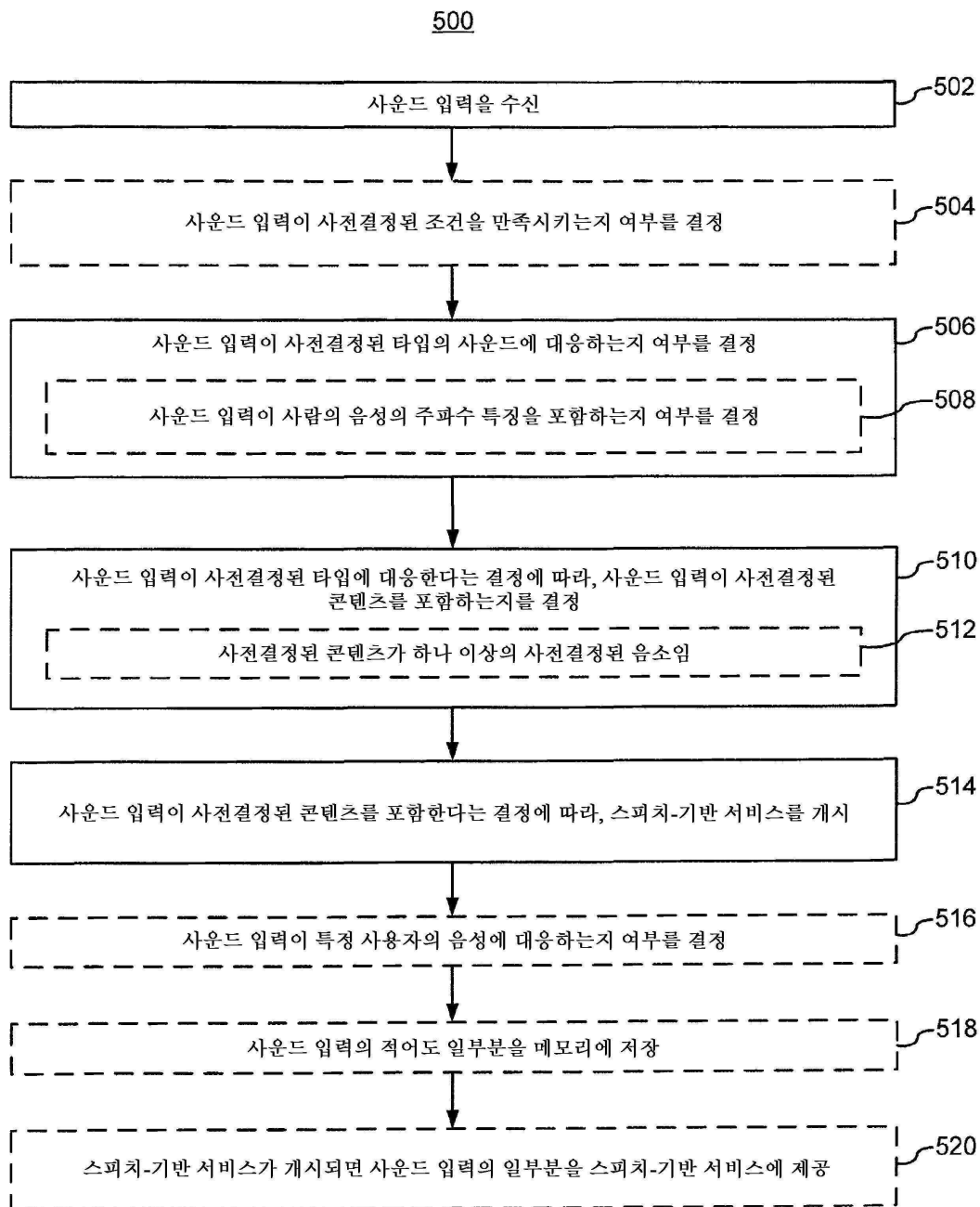
도면3c



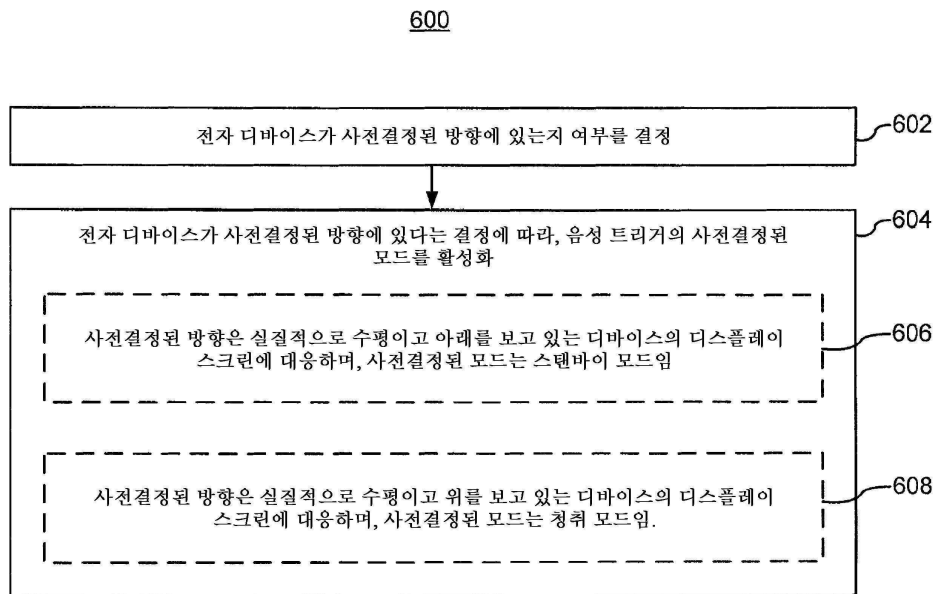
도면4



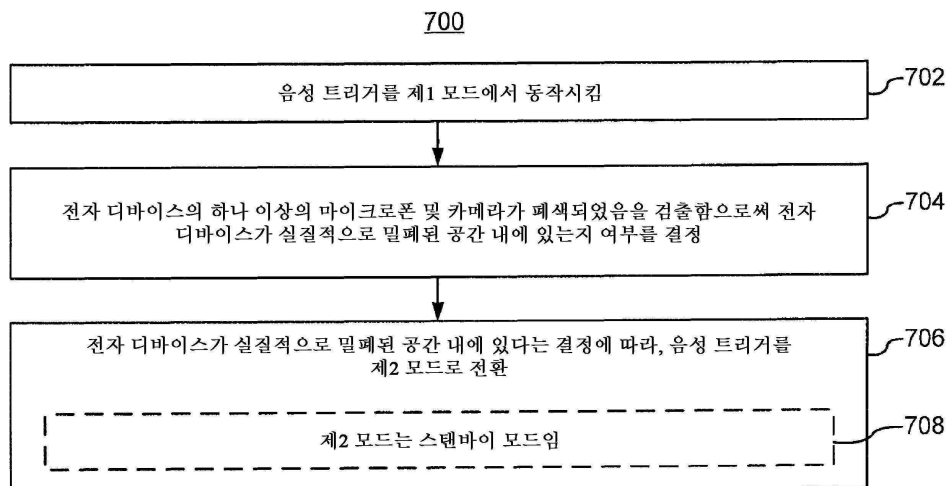
도면5



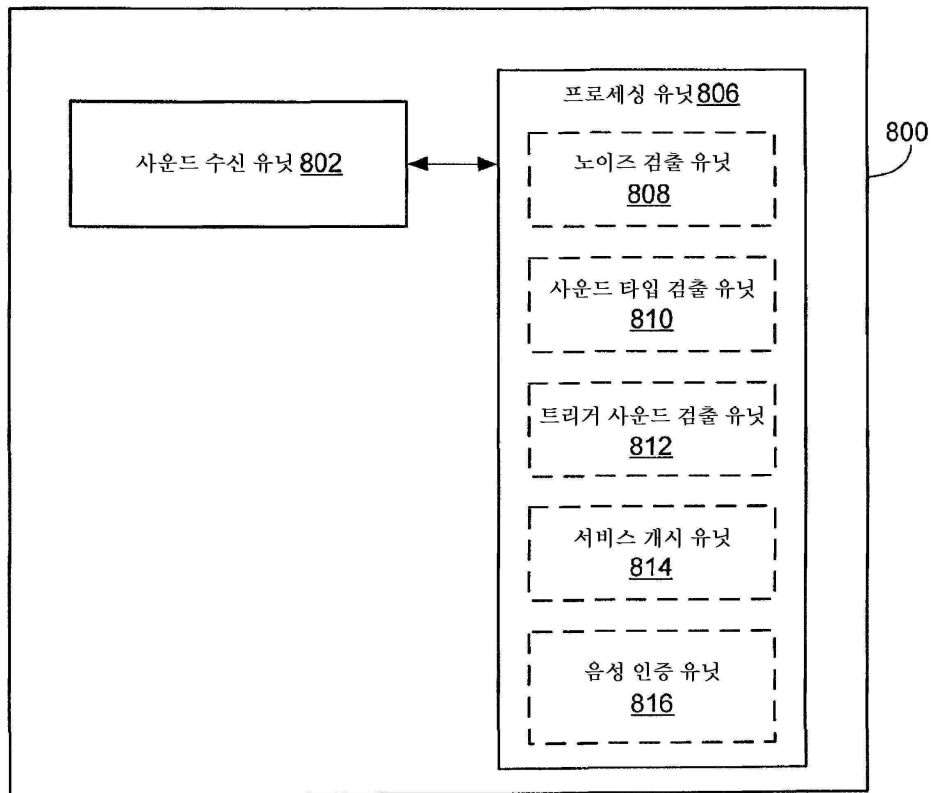
도면6



도면7



도면8



도면9

