



**(19) 대한민국특허청(KR)**  
**(12) 등록특허공보(B1)**

(45) 공고일자 2022년12월12일  
(11) 등록번호 10-2476621  
(24) 등록일자 2022년12월07일

- |  |  |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/>G10L 15/22 (2006.01) G06F 3/16 (2018.01)<br/>G06F 9/448 (2018.01) G10L 13/027 (2013.01)<br/>G10L 13/08 (2006.01)</p> <p>(52) CPC특허분류<br/>G10L 15/22 (2013.01)<br/>G06F 3/167 (2013.01)</p> <p>(21) 출원번호 10-2020-7033522</p> <p>(22) 출원일자(국제) 2018년05월07일<br/>심사청구일자 2020년11월20일</p> <p>(85) 번역문제출일자 2020년11월20일</p> <p>(65) 공개번호 10-2021-0002599</p> <p>(43) 공개일자 2021년01월08일</p> <p>(86) 국제출원번호 PCT/US2018/031444</p> <p>(87) 국제공개번호 WO 2019/216875<br/>국제공개일자 2019년11월14일</p> <p>(56) 선행기술조사문헌<br/>KR1020160003138 A*<br/>WO2011088053 A2*<br/>*는 심사관에 의하여 인용된 문헌</p> | <p>(73) 특허권자<br/>구글 엘엘씨<br/>미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)</p> <p>(72) 발명자<br/>키라즈시 올라스<br/>미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)<br/>코임브라 아담<br/>미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/>박장원</p> |
|--|--|

전체 청구항 수 : 총 14 항

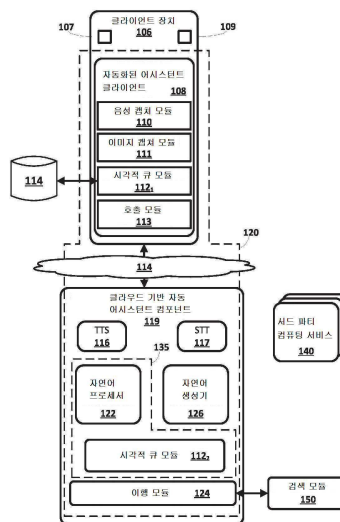
심사관 : 안지현

**(54) 발명의 명칭 사용자, 자동화된 어시스턴트 및 컴퓨팅 서비스 간의 다중 모드 상호 작용**

**(57) 요약**

사용자, 자동화된 어시스턴트 및 기타 컴퓨팅 서비스 간의 다중 모드 상호 작용에 대한 기술이 여기에서 설명된다. 다양한 구현에서, 사용자는 제3자(서드 파티) 컴퓨팅 서비스에 더 참여하기 위해 자동화된 어시스턴트와 참여할 수 있다. 일부 구현에서, 사용자는 시각적/촉각 양식과 같은 구두 양식 이외의 입력 양식 및 구두 입력 양식 둘 다를 사용하여 제3자 컴퓨팅 서비스와 관련된 대화 상태 기계를 통해 진행할 수 있다.

**대표도 - 도1**



(52) CPC특허분류

*G06F 9/4498* (2018.02)

*G10L 13/027* (2013.01)

*G10L 13/08* (2013.01)

*G10L 2015/223* (2013.01)

*G10L 2015/228* (2013.01)

(72) 발명자

**이 아브라함**

미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이  
1600 (우:94043)

**동 웨이**

미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이  
1600 (우:94043)

**아마라시리와의르테나 투산**

미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이  
1600 (우:94043)

## 명세서

### 청구범위

#### 청구항 1

시스템으로서,

디스플레이부를 구비하는 적어도 하나의 클라이언트 컴퓨팅 장치 - 상기 클라이언트 컴퓨팅 장치는 자동화된 어시스턴트(assistant)의 클라이언트부(client portion)를 구현함 -;

상기 자동화된 어시스턴트의 서버부(server portion)를 구현하는 적어도 하나의 서버 컴퓨팅 장치를 포함하도록 형성되며,

사용자는, 사용자, 자동화된 어시스턴트 및 서드 파티(third party) 컴퓨팅 서비스 사이의 인간 대 컴퓨터 대화 세션에 참여하기 위해 상기 클라이언트부와 상호 작용하고;

상기 서드 파티 컴퓨팅 서비스는, 구두 대화 상태 머신(verbal dialog state machine) 및 시각적 대화 상태 머신(visual dialog state machine)에 따라 사용자와 참여하며;

상기 자동화된 어시스턴트의 클라이언트부는,

상기 자동화된 어시스턴트의 서버부를 통해, 상기 서드 파티 컴퓨팅 서비스로부터 상기 시각적 대화 상태 머신의 제1 시각적 대화 상태를 나타내는 데이터를 수신하고,

상기 제1 시각적 대화 상태에 기초하여, 상기 인간 대 컴퓨터 대화 세션과 관련된 그래픽 사용자 인터페이스를 디스플레이부에 렌더링하며, - 상기 그래픽 사용자 인터페이스는, 상기 구두 대화 상태 머신이, 상기 제1 시각적 대화 상태에 대응하는 제1 구두 대화 상태에서 제2 구두 대화 상태로 전환되도록 하는 적어도 하나의 그래픽 요소를 포함함 - ;

사용자에 의한 적어도 하나의 그래픽 요소의 작동을 감지하고,

적어도 하나의 그래픽 요소의 작동을 나타내는 데이터를 자동화된 어시스턴트의 서버부에 제공하도록 구성되며,

상기 자동화된 어시스턴트의 서버부는, 상기 적어도 하나의 그래픽 요소의 작동을 나타내는 데이터를 상기 서드 파티 컴퓨팅 서비스에 제공하도록 구성되며,

상기 적어도 하나의 그래픽 요소의 작동을 나타내는 데이터는, 상기 서드 파티 컴퓨팅 서비스가, 상기 제1 구두 대화 상태에서 상기 제2 구두 대화 상태로 전환되도록 하는 것을 특징으로 하는 시스템.

#### 청구항 2

제1항에 있어서, 상기 자동화된 어시스턴트의 서버부는,

상기 서드 파티 컴퓨팅 서비스로부터 상기 제2 구두 대화 상태를 나타내는 데이터를 수신하고,

상기 자동화된 어시스턴트의 클라이언트부에 상기 제2 구두 대화 상태를 나타내는 데이터를 제공하도록 구성되는 것을 특징으로 하는 시스템.

#### 청구항 3

제1항에 있어서,

상기 제2 구두 대화 상태를 나타내는 데이터는, 텍스트 또는 오디오 데이터를 포함하고,

상기 클라이언트 컴퓨팅 장치는, 상기 제2 구두 대화 상태를 표시하는 데이터를 수신하는 것에 응답하여, 상기 텍스트를 음성으로 변환하고 상기 음성을 가청음으로 렌더링하거나, 또는 오디오 데이터를 가청음으로 렌더링하도록 더 구성되는 것을 특징으로 하는 시스템.

#### 청구항 4

제1항에 있어서,

상기 자동화된 어시스턴트의 서버부로부터, 상기 서드 파티 컴퓨팅 서비스로의 상기 적어도 하나의 그래픽 요소의 작동을 나타내는 데이터의 제공에 의하여, 상기 서드 파티 컴퓨팅 서비스가 상기 제1 시각적 대화 상태에서부터 제2 시각적 대화 상태로 전환하는 것을 특징으로 하는 시스템.

**청구항 5**

제4항에 있어서, 상기 제2 시각적 대화 상태는,

상기 제2 구두 대화 상태에 대응하는 것을 특징으로 하는 시스템.

**청구항 6**

제5항에 있어서, 상기 자동화된 어시스턴트의 클라이언트부는,

상기 자동화된 어시스턴트의 서버부를 통해 상기 서드 파티 컴퓨팅 서비스로부터 상기 제2 시각적 대화 상태를 나타내는 데이터를 수신하고,

상기 제2 시각적 대화 상태에 기반하여, 상기 인간 대 컴퓨터 대화 세션과 관련된 업데이트된 그래픽 사용자 인터페이스를 상기 디스플레이부에 렌더링하도록 더 구성되는 것을 특징으로 하는 시스템.

**청구항 7**

제6항에 있어서, 상기 제2 시각적 대화 상태를 나타내는 데이터는,

상기 클라이언트 컴퓨팅 장치가, 상기 업데이트된 그래픽 사용자 인터페이스를 렌더링하기 위해 비접촉 상호 작용을 자동으로 수행하게 하는 것을 특징으로 하는 시스템.

**청구항 8**

제7항에 있어서, 상기 비접촉 상호 작용은,

상기 그래픽 사용자 인터페이스에서 선택 가능한 요소의 작동, 상기 그래픽 사용자 인터페이스에서 렌더링된 문서를 특정 위치로 스크롤, 및 상기 그래픽 사용자 인터페이스의 일부 확대 중, 하나 이상을 포함하는 것을 특징으로 하는 시스템.

**청구항 9**

제1항에 있어서,

상기 서드 파티 컴퓨팅 서비스가 상기 제1 구두 대화 상태에서 상기 제2 구두 대화 상태로 전환된 이후, 상기 자동화된 어시스턴트의 클라이언트부는,

상기 클라이언트 컴퓨팅 장치의 하나 이상의 입력 컴포넌트를 통해 사용자로부터 음성 자유 형식 입력(vocal free form input)을 수신하고,

상기 자동화된 어시스턴트의 서버부에 상기 음성 자유 형식 입력을 나타내는 데이터를 제공하며,

상기 자동화된 어시스턴트의 서버부로부터 디스플레이 컨텍스트(context)를 나타내는 데이터를 수신하고 - 상기 자동화된 어시스턴트의 서버부는, 음성 자유 형식 입력을 나타내는 데이터 또는 음성 자유 형식 입력을 나타내는 데이터를 기반으로 하는 추가 데이터를 상기 서드 파티 컴퓨팅 서비스에 제공하는 것을 기반으로 상기 서드 파티 컴퓨팅 서비스로부터 디스플레이 컨텍스트를 나타내는 데이터를 획득하고; 상기 서드 파티 컴퓨팅 서비스는, 상기 수신된 데이터 또는 상기 추가 데이터에 기초하고, 그리고 상기 제2 구두 대화 상태로의 전환에 기초하여 디스플레이 컨텍스트를 나타내는 데이터를 생성함 -,

수신된 디스플레이 컨텍스트를 기반으로, 상기 그래픽 사용자 인터페이스에서 비접촉(touchless) 상호 작용을 트리거하도록 구성되는 것을 특징으로 하는 시스템.

**청구항 10**

제9항에 있어서,

상기 비접촉 상호 작용은,

상기 그래픽 사용자 인터페이스에서 선택 가능한 요소의 작동, 상기 그래픽 사용자 인터페이스에서 렌더링된 문서를 특정 위치로 스크롤, 및 상기 그래픽 사용자 인터페이스의 일부 확대 중, 하나 이상을 포함하는 것을 특징으로 하는 시스템.

**청구항 11**

시스템으로서,

디스플레이부를 구비하는 적어도 하나의 클라이언트 컴퓨팅 장치 - 상기 클라이언트 컴퓨팅 장치는 자동화된 어시스턴트(assistant)의 클라이언트부(client portion)를 구현함 -;

상기 자동화된 어시스턴트의 서버부(server portion)를 구현하는 적어도 하나의 서버 컴퓨팅 장치를 포함하도록 구성되며,

사용자는, 사용자, 자동화된 어시스턴트 및 서드 파티(third party) 컴퓨팅 서비스 사이의 인간 대 컴퓨터 대화 세션에 참여하기 위해 상기 클라이언트부와 상호 작용하고;

상기 자동화된 어시스턴트의 클라이언트부는,

상기 클라이언트 컴퓨팅 장치의 디스플레이부 상에 상기 인간 대 컴퓨터 대화 세션과 관련된 그래픽 사용자 인터페이스를 렌더링하고,

상기 클라이언트 컴퓨팅 장치의 하나 이상의 입력 컴포넌트에서 사용자로부터 음성 자유 형식 입력을 수신하고 -상기 음성 자유 형식 입력은 상기 사용자에게 의해 공식화됨(formulated)-,

상기 자동화된 어시스턴트의 서버부에 상기 음성 자유 형식 입력을 나타내는 데이터를 제공하고,

상기 자동화된 어시스턴트의 서버부에서 디스플레이 콘텐츠를 나타내는 데이터를 수신하고,

상기 디스플레이 콘텐츠에 기초하여 사용자와 그래픽 사용자 인터페이스 간의 비접촉 상호 작용을 트리거하도록 구성되며,

상기 자동화된 어시스턴트의 서버부는,

상기 음성 자유 형식 입력을 나타내는 데이터를 기반으로 사용자의 의도를 결정하고,

상기 서드 파티 컴퓨팅 서비스에 의도를 나타내는 데이터를 제공하고 - 상기 의도를 나타내는 데이터는 상기 서드 파티 컴퓨팅 서비스가 해결 정보를 생성하도록 의도를 해결하도록 함 -,

상기 의도를 나타내는 데이터의 제공에 대한 응답으로, 상기 서드 파티 컴퓨팅 서비스로부터 상기 디스플레이 콘텐츠를 수신하도록 구성되며,

상기 디스플레이 콘텐츠는 상기 인간 대 컴퓨터 대화 세션과 관련된 상기 서드 파티 컴퓨팅 서비스를 위해 유지되고, 상기 사용자의 의도 및 상기 해결 정보 중 적어도 하나에 부분적으로 기초하거나 상기 사용자의 의도 및 상기 해결 정보 모두에 기초하여 업데이트되는 것을 특징으로 하는 시스템.

**청구항 12**

제11항에 있어서, 상기 그래픽 사용자 인터페이스는,

상기 자동화된 어시스턴트의 클라이언트부에 내장된 웹 브라우저를 포함하는 것을 특징으로 하는 시스템.

**청구항 13**

제12항에 있어서, 상기 비접촉 상호 작용은,

상기 그래픽 사용자 인터페이스에서 선택 가능한 요소의 작동, 상기 그래픽 사용자 인터페이스에서 렌더링된 문서를 특정 위치로 스크롤, 및 상기 그래픽 사용자 인터페이스의 일부 확대 중, 하나 이상을 포함하는 것을 특징으로 하는 시스템.

**청구항 14**

제11항에 있어서, 상기 음성 자유 형식 입력은 상기 사용자에게 의한 선택을 위해 제공된 옵션들의 그룹으로 제한되지 않는 것을 특징으로 하는 시스템.

**청구항 15**

삭제

**청구항 16**

삭제

**청구항 17**

삭제

**청구항 18**

삭제

**청구항 19**

삭제

**청구항 20**

삭제

**청구항 21**

삭제

**청구항 22**

삭제

**청구항 23**

삭제

**발명의 설명**

**기술 분야**

[0001] 본 발명은 자동화된 어시스턴트(assistant)를 통해 사용자와 컴퓨팅 서비스 간에 다양한 모드의 상호 작용을 가능하게 하는 시스템에 대한 것이다.

**배경 기술**

[0002] 인간은 "자동화된 어시스턴트"("채팅 봇", "대화형 개인 비서", "지능형 개인 비서", "개인 음성 비서"등 라고도 함)라고 하는 대화형 소프트웨어 응용 프로그램을 사용하여 인간 대 컴퓨터 대화를 소통할 수 있다. 예를 들어, 인간은 다음을 포함 할 수 있는 자유 형식 자연어 입력을 사용하여 명령, 질의 및/또는 요청(통칭하여 "쿼리"라고 함)을 제공 할 수 있다.

[0003] 예를 들어, 인간(이하 상기 자동화된 어시스턴트와 상호 작용할 때 "사용자"라고 함)은 텍스트로 변환 된 음성 발화를 포함할 수 있는 자유 형식 자연어 입력을 사용하여 명령, 쿼리 및/또는 요청(통칭하여 질의라고 함)을 제공한 다음, 처리 및/또는 형식화된 자유 형식 자연어 입력을 제공 할 수 있다. 대부분의 경우 자동화된 어시스턴트는 사전 정의된 구두 호출 구문을 사용하여 먼저 "호출(invoked)"되어야 한다.

[0004] 한편 자동화된 어시스턴트와 상호 작용할 수 있는 많은 컴퓨팅 서비스("소프트웨어 에이전트"또는 "에이전트"라고도 함)가 존재한다. 이러한 컴퓨팅 서비스를 제공하는 주체는 종종 자동화된 어시스턴트를 제공하는 주체와 직접 관련이 없기 때문에, 이러한 컴퓨팅 서비스는 종종 "서드 파티(third party)"(또는 "서드 파티 개발자")라

는 개발자가 개발 및/또는 제공한다.

- [0005] 그러나 컴퓨팅 서비스는 서드 파티가 개발한 서비스에 국한되지 않으며, 자동화된 어시스턴트를 구현하는 동일한 엔티티(entity)에 의해 구현 될 수 있다. 컴퓨팅 서비스는 다양한 사용자 의도를 해결하도록 구성될 수 있으며, 이들 중 대부분은 자동화된 지원으로 해결되지 않을 수 있다.
- [0006] 이러한 의도는 스마트 장치 제어 또는 구성, 작업 수행을 위한 단계별 지침 수신, 온라인 서비스와의 상호 작용에 관련될 수 있지만 이에 국한되지는 않는다.
- [0007] 따라서 많은 자동화된 어시스턴트는 사용자 및 서드 파티 컴퓨팅 서비스 모두와 동시에 상호 작용하여, 효과적으로 사용자와 서드 파티 사이의 중개자 또는 중개자 역할을 한다.
- [0008] 일부 서드 파티 컴퓨팅 서비스들은 사용자 및/또는 센서나 웹 서비스 등으로부터 수신된 다양한 입력에 기초하여 발생하는 복수의 상태뿐만 아니라 이러한 상태 사이의 전환을 효과적으로 정의하는 대화 상태 머신들(dialog state machines)에 따라 작동할 수 있다.
- [0009] 사용자와 상기 서드 파티 컴퓨팅 서비스에서 하나 이상의 대화가 "순환되는(turns)"동안, 사용자가 (중개자로서 자동화된 어시스턴트를 통해) 자유로운 형식의 자연어 입력(음성 또는 입력)을 제공할 때, 서드 파티 컴퓨팅 서비스와 관련된 대화 상태 머신은, 다양한 대화 상태 사이를 진행하여 결국 사용자의 의도가 해결되는 상태에 도달할 수 있다.

**발명의 내용**

- [0010] 자동화된 어시스턴트가 더욱 보편화됨에 따라, 자동화된 어시스턴트와의 상호 작용을 용이하게 하도록 특별히 설계된 컴퓨팅 장치(이하 "보조 기기(assistant device)"라고 함)가 더욱 보편화되고 있다. 많은 보조 기기를 통해 사용자는 자동화된 어시스턴트와 터치 없이 상호 작용할 수 있다. 예를 들어, 보조 기기에는 사용자가 음성 발화를 입력으로 제공 할 수 있는 마이크가 포함되는 경우가 있다. 또한 점점 더 많은 보조 장치에 디스플레이 기능이 포함될 수 있다.
- [0011] 보다 구체적으로, 배타적이지는 않지만, 본 명세서에 설명된 일부 구현은 사용자가 시각/촉각 양식과 같은 언어적 자유 형식 자연어 입력 이외의 입력 양식을 사용하여 서드 파티 컴퓨팅 서비스와 관련된 대화 상태 머신에 따라 진행할 수 있게 한다. 본 명세서에서 사용되는 바와 같이, "구두"는 키보드와 상호 작용 및/또는 음성 발화(음성 인식을 사용하여 텍스트 콘텐츠로 변환될 수 있음)를 제공함으로써 사용자에게 의해 제공 될 수 있는 자유 형식의 자연어 입력을 의미할 것이다. 추가적으로 또는 대안적으로, 여기에 설명된 일부 구현은 사용자가 서드 파티 컴퓨팅 서비스와 관련된 그래픽 사용자 인터페이스와 비접촉(touchless) 상호 작용을 트리거 할 수 있게 한다. 본 명세서에서 다른 변형이 고려된다. 사용자가 언어 및 비언어적 입력의 조합을 사용하여 서드 파티 서비스와 상호 작용할 수 있도록 하는 메커니즘을 제공하면 신체 기능이 있는 사용자와 광범위한 장애를 가진 사용자 모두 서드 파티 서비스에 대한 액세스를 용이하게 할 수 있다(특별하게 설정된 버전의 제공에 대한 필요성이 감소할 수 있다). 이는 메커니즘을 통해 사용자가 자신에게 가장 적합한 방식으로 서드 파티 서비스와 상호 작용할 수 있기 때문이다.
- [0012] 일부의 구현 예에서, 자동화된 어시스턴트는 클라이언트부(client portion)와 서버부(server portion)를 포함할 수 있다. 자동화된 어시스턴트의 클라이언트부는 클라이언트 장치(예 : 스마트 폰, 스마트 워치, 보조 장치 등)에서 "보조 앱"으로 작동할 수 있다. 자동화된 어시스턴트의 서버부는 본 명세서에서 때때로 "클라우드"컴퓨팅 시스템 또는 "클라우드"로 지칭되는 하나 이상의 서버에서 작동할 수 있다. 다양한 구현에서, 클라이언트 장치는 사용되는 디스플레이부를 포함할 수 있다. 그래픽 사용자 인터페이스("GUI")를 렌더링하기 위해 자동화된 어시스턴트의 클라이언트부에 의해. 일부 구현에서, GUI는 서드 파티 컴퓨팅 서비스에 의해 제공되는 기본 마크업 언어(Markup language)(예를 들어, HTML, XML)에 기초하여 그래픽을 렌더링하는 임베디드 웹 브라우저의 형태를 가질 수 있다. GUI의 제공은 사용자가 오디오 상호 작용만을 통해 가능한 것보다 더 효율적인 방식으로 서드 파티 컴퓨팅 서비스와 상호 작용할 수 있게 한다.
- [0013] 사용자가 서드 파티 컴퓨팅 서비스에 더 참여하기 위해 자동화된 어시스턴트와 참여하는 일부의 구현 예에서, 서드 파티 컴퓨팅 서비스는 구두 대화 상태 머신 및/또는 시각적 대화 상태 머신에 따라 사용자와 참여할 수 있다. 구두 대화 상태 머신은, 사용자와, 사용자가 타이핑 및/또는 음성 자유 형식 자연어 입력을 제공하는 컴퓨팅 서비스 사이의 음성 대화를 관리 할 수 있다. 시각적 대화 상태 시스템은 몰입형 경험의 일부로 GUI에서 렌더링되는 것을 제어할 수 있다. 이러한 일부 구현 예에서, 구두 대화 상태 머신의 하나 이상의 구두 대화 상태

는, 구두 대화가 특정 상태에 도달할 때 특정 GUI가 렌더링되도록 시각적 대화 상태 머신의 하나 이상의 시각적 대화 상태에 링크될 수 있다. 그리고/또는 사용자가 시각적 대화 상태 머신을 특정 시각적 대화 상태로 진행하기 위해 특정 방식으로 GUI와 상호 작용할 때, 상기 구두 대화는 자동으로 특정 상태로 진행될 수 있다.

[0014] 일부의 구현 예에서, 사용자는 자동화된 어시스턴트의 서버부를 통해 유입되는 구두 명령 또는 기타 비접촉 입력(사용자의 의도를 나타내는)을 사용하여 서드 파티 컴퓨팅 서비스에 의해 제공되는 GUI와 비접촉식으로 상호 작용할 수 있다. 예를 들어, 사용자가 스마트 홈 장치를 제어 할 수 있는 서드 파티 컴퓨팅 서비스를 사용(자동화된 어시스턴트를 통해)한다고 가정할 수 있다. 이 경우 스마트 홈 장치를 제어/구성하는 것과 관련된 다양한 옵션이 GUI 상에 표시( 예를 들어 복수의 GUI 타일들)될 수 있다. 사용자는 타일들 중 하나를 선택(예 : 마우스 또는 터치 스크린 사용)할 수 있을 뿐만 아니라, "내 장치 구성(Let's configure my device)"또는 "도움이 필요합니다(I need help)."와 같은 구두 명령을 출력하기 위한 타일들 중 하나를 선택할 수도 있다. 이는, 사용자가 장치를 구성할 수 있도록 하거나 선택 가능한 FAQ 목록을 제공하는 인터페이스 모듈로 진행되는 GUI와의 비접촉 상호 작용을 트리거(trigger) 할 수 있다. 또 다른 예로, 사용자는 "한 화면 스크롤 다운(scroll down one screen)"또는 "다음 단락으로 이동(move to the next paragraph)"과 같은 구두 명령을 실행하여 GUI의 일부로 렌더링 된 문서를 스크롤 할 수 있다. 자동화된 어시스턴트의 서버부를 통해 서드 파티 컴퓨팅 서비스에 구두 명령/기타 비접촉 입력을 입력하는 것 외에도, 서드 파티 컴퓨팅 서비스에서 클라이언트 장치로 보내는 모든 데이터 및 명령은 자동화된 어시스턴트의 서버부를 통해 전달될 수 있다. 이것은 자동화된 어시스턴트의 서버부가, 서드 파티 컴퓨팅 서비스로부터 발생하거나 상기 서드 파티 컴퓨팅 서비스를 통해 수신되는 불법적인 메시지 등이 클라이언트 장치에 도달하는 것을 방지할 수 있기 때문에, 클라이언트 장치의 보안을 증가시키는 역할을 할 수 있다.

[0015] 일부의 구현 예에서, 사용자는 시각적 대화 상태 머신과 구두 대화 상태 머신 모두에 따라 진행하기 위해, 서드 파티 컴퓨팅 서비스에 의해 제공되는 GUI와 상호 작용할 수 있다. 예를 들어, 사용자는 GUI를 조작하여 특정 객체를 확대한 다음 "What 's this?"와 같은 구두 질의를 제공할 수 있다. GUI를 특정 객체에 집중함으로써 사용자는 구두 상태 머신을 특정 객체에 초점을 맞춘 상태로 진행시킬 수 있다. 결과적으로 "What 's this?"라는 질의에서 "this"라는 용어는 특정 개체로 해결될 수 있다.

[0016] 다른 예로서, 서드 파티 컴퓨팅 서비스가 다중 선택 차례(turn) 기반 대화(예를 들어, 작업 수행을 위한 단계별 안내와 관련)를 제공한다고 가정할 수 있다. 특정 질문에 대해, 클라이언트 컴퓨팅 장치에서 렌더링되는 GUI는 서드 파티 컴퓨팅 서비스의 시각적 대화 상태 머신의 특정 상태에 대응할 수 있으며, 객관식 질문 및 답변 선택을 포함할 수 있다. 동시에, 서드 파티 컴퓨팅 서비스의 구두 대화 상태 머신에 따라 구두 자연어 출력이 제공될 수 있다. 예를 들어, 언어적 자연어 출력은 객관식 질문을 소리내어 읽을 수 있다(그리고 어떤 경우에는 답변 선택도 진행할 수 있다). 일부의 구현 예에서, 사용자는 화면상의 답변 선택 중 하나를 선택하거나 특정 답변 선택을 나타내는 음성 발화를 제공함으로써, 시각적 및, 구두 대화 상태 머신 모두를 상이한 각각의 상태(예를 들어 다음 질문)로 진행시킬 수 있다.

[0017] 여기에 설명된 기술은 다양한 기술적 이점 및 혜택을 야기할 수 있다. 예를 들어, 자유 형식 입력을 기반으로 의도 일치 해석 및/또는 수행하는 것은 상당한 컴퓨팅 리소스를 필요로 한다. 그러나 사용자가 사전 정의된 그래픽 요소와 촉각적으로 상호 작용하여 자동화된 어시스턴트와 대화를 진행할 수 있도록 함으로써 해석/의도 일치를 수행할 필요가 없고, 서드 파티 컴퓨팅 서비스와 관련된 대화 상태 머신의 효율적인 탐색을 가능하게 함으로써 이러한 리소스를 절약할 수 있다. 이러한 낭비 또는 전력 및 컴퓨팅 리소스의 비효율적이거나 불필요한 사용은, 클라이언트 컴퓨팅 장치 자체 및/또는 자동화된 어시스턴트의 서버부를 운영하는 하나 이상의 네트워크 서버와 같은 원격 컴퓨팅 장치에서 발생할 수 있다. 또한 원격 컴퓨팅 장치와의 불필요한 통신은 통신 네트워크에 원치 않는 부하를 유발한다. 또 다른 예로, 플랫폼에 구애받지 않는 웹 브라우저를 자동화된 어시스턴트의 클라이언트부에 내장함으로써, 서드 파티 개발자는 일관된 방식으로 컴퓨팅 서비스에 수반되는 시각적 콘텐츠를 제공할 수 있다. 이것은 특정 클라이언트 장치에 대해 특별히 구성되는 메커니즘/프로세스를 요구하지 않고 컴퓨팅 서비스 및 시각적 콘텐츠가 광범위한 클라이언트 장치에 제공 될 수 있도록 한다.

[0018] 일부의 구현 예에서, 적어도 하나의 클라이언트 컴퓨팅 장치를 포함하는 시스템이 제공된다. 클라이언트 컴퓨팅 장치는 디스플레이부를 가지며 자동화된 어시스턴트의 클라이언트부를 구현한다. 시스템은 자동화된 어시스턴트의 서버부를 구현하는 하나 이상의 서버 컴퓨팅 장치를 더 포함한다. 사용자는 클라이언트부와 상호 작용하여 사용자, 자동화된 어시스턴트 및 서드 파티 컴퓨팅 서비스 간의 인간 대 컴퓨터 대화 세션에 참여한다. 서드 파티 컴퓨팅 서비스는 구두 대화 상태 머신 및 시각적 대화 상태 머신에 따라 사용자와 연결된다. 자동화된 어시스턴트의 클라이언트부는 자동화된 어시스턴트의 서버부를 통해 서드 파티 컴퓨팅 서비스로부터 시각적 대화 상



태 머신의 제1 시각적 대화 상태를 나타내는 데이터를 수신하도록 구성된다. 자동화된 어시스턴트의 클라이언트부는 또한 인간 대 컴퓨터 대화 세션과 연관된 그래픽 사용자 인터페이스를 제1 시각적 대화 상태에 기초하여 디스플레이부에 렌더링하도록 구성된다. 그래픽 사용자 인터페이스는 언어 상태 머신이 제1 시각적 대화 상태에 대응하는 제1 구두 대화 상태에서 제2 구두 대화 상태로 전환시키는 적어도 하나의 그래픽 요소를 포함한다. 자동화된 어시스턴트의 클라이언트부는 사용자에게 의한 적어도 하나의 그래픽 요소의 동작을 검출하도록 추가로 구성된다. 또한 자동화된 어시스턴트의 클라이언트부는 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터를 자동화된 어시스턴트의 서버부에 제공하도록 구성된다. 자동화된 어시스턴트의 서버부는 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터를 서드 파티 컴퓨팅 서비스에 제공하도록 구성된다. 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터는 서드 파티 컴퓨팅 서비스가 제1 구두 대화 상태에서 제2 구두 대화 상태로 전환된다.

- [0019] 여기에 개시된 기술의 이들 및 다른 구현은 선택적으로 다음 특징 중 하나 이상을 포함할 수 있다.
- [0020] 일부의 구현 예에서, 일부의 구현 예에서, 자동화된 어시스턴트의 서버부는 서드 파티 컴퓨팅 서비스로부터 제2 구두 대화 상태를 나타내는 데이터를 수신하고, 자동화된 어시스턴트의 클라이언트부에 제2 구두 대화 상태를 나타내는 데이터를 제공하도록 추가로 구성될 수 있다. 일부의 구현 예에서, 제2 구두 대화 상태를 나타내는 데이터는 텍스트 또는 오디오 데이터를 포함할 수 있다. 이들 구현 중 일부에서, 클라이언트 장치는 제2 구두 대화 상태를 나타내는 데이터를 수신하는 것에 응답하여 텍스트를 음성으로 변환하고 음성을 청각 적으로 렌더링하거나 오디오 데이터를 청각 적으로 렌더링하도록 추가로 구성 될 수 있다.
- [0021] 일부의 구현 예에서, 자동화된 어시스턴트의 서버부로부터 서드 파티 컴퓨팅 서비스로의 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터의 제공은 서드 파티 컴퓨팅 서비스가 제1 시각적 대화 상태에서 제2 시각적 대화 상태로 전환되도록 할 수 있다. 이러한 구현 중 일부에서는, 자동화된 어시스턴트의 클라이언트부가, 자동화된 어시스턴트의 서버부를 통해 서드 파티 컴퓨팅 서비스로부터 제2 시각적 대화 상태를 나타내는 데이터를 수신하고, 인간 대 컴퓨터 대화 세션과 관련된 업데이트된 그래픽 사용자 인터페이스를 제2 시각적 대화 상태에 기반하여 디스플레이부에 렌더링하도록 추가로 구성될 수 있다. 이러한 구현 중 일부에서, 제2 시각적 대화 상태를 나타내는 데이터는 클라이언트 장치가 업데이트된 그래픽 사용자 인터페이스를 렌더링하기 위해, 비접촉 상호 작용을 자동으로 수행하도록 할 수 있다. 이러한 구현 중 일부에서 비접촉 상호 작용은, 그래픽 사용자 인터페이스의 선택 가능한 요소 작동, 그래픽 사용자 인터페이스에서 렌더링된 문서의 특정 위치로 스크롤, 및 그래픽 사용자 인터페이스의 일부 확대 중 하나 이상을 포함할 수 있다.
- [0022] 일부의 구현 예에서, 서드 파티 컴퓨팅 서비스가 제1 구두 대화 상태에서 제2 구두 대화 상태로 전이 한 후, 자동화된 어시스턴트의 클라이언트 부는 클라이언트 컴퓨팅 장치의 하나 이상의 입력 컴포넌트에서 사용자로부터 음성 자유 형식 입력을 수신하고, 자동 어시스턴트의 서버부에 음성 자유 형식 입력을 나타내는 데이터를 제공하며, 자동화된 어시스턴트의 서버부로부터 디스플레이 컨텍스트를 나타내는 데이터를 수신하도록 추가로 구성될 수 있다. 자동화된 어시스턴트의 서버부는, 음성 자유 형식 입력을 나타내는 데이터 또는 음성 자유 형식 입력을 나타내는 데이터를 기반으로 하는 추가 데이터를 서드 파티 컴퓨팅 서비스에 제공하는 것에 기초하여, 서드 파티 컴퓨팅 서비스에서 디스플레이 컨텍스트를 나타내는 데이터를 획득할 수 있다. 서드 파티 컴퓨팅 서비스는 수신된 데이터 또는 추가 데이터에 기초하고, 제2 구두 대화 상태로의 전이에 기초하여, 디스플레이 컨텍스트를 나타내는 데이터를 생성할 수 있다. 자동화된 어시스턴트의 클라이언트부는 수신된 디스플레이 컨텍스트에 기초하여 그래픽 사용자 인터페이스에서 비접촉 상호 작용을 트리거하도록 추가로 구성될 수 있다.
- [0023] 일부의 구현 예에서, 비접촉 상호 작용은 그래픽 사용자 인터페이스의 선택 가능한 요소의 작동, 그래픽 사용자 인터페이스에서 렌더링된 문서의 특정 위치로 스크롤, 그래픽 사용자 인터페이스의 일부 확대 중 하나 이상을 포함할 수 있다.
- [0024] 일부의 구현 예에서, 적어도 하나의 클라이언트 컴퓨팅 장치 및 하나 이상의 서버 컴퓨팅 장치를 포함하는 시스템이 제공된다. 클라이언트 컴퓨팅 장치는 디스플레이부를 가지며 자동화된 어시스턴트의 클라이언트부를 구현한다. 시스템은 자동화된 어시스턴트의 서버부를 구현하는 하나 이상의 서버 컴퓨팅 장치를 더 포함한다. 사용자는 클라이언트부와 상호 작용하여 사용자, 자동화된 어시스턴트 및 서드 파티 컴퓨팅 서비스 간의 인간 대 컴퓨터 대화 세션에 참여한다. 자동화된 어시스턴트의 클라이언트부는 인간 대 컴퓨터 대화 세션과 연관된 그래픽 사용자 인터페이스를 클라이언트 컴퓨팅 장치의 디스플레이부에 렌더링하도록 구성된다. 자동화된 어시스턴트의 클라이언트부는 또한 클라이언트 컴퓨팅 장치의 하나 이상의 입력 컴포넌트에서 사용자로부터 음성 자유 형식 입력을 수신하도록 구성된다. 자동화된 어시스턴트의 클라이언트부는 또한 자동 어시스턴트의 서버부에 보컬 자

유 형식 입력을 나타내는 데이터를 제공하도록 구성된다. 자동화된 어시스턴트의 클라이언트부는 또한 디스플레이 컨텍스트를 나타내는 데이터를 자동화된 어시스턴트의 서버부로부터 수신하도록 구성된다. 자동화된 어시스턴트의 클라이언트부는 디스플레이 컨텍스트에 기초하여 사용자와 그래픽 사용자 인터페이스 간의 비접촉 상호 작용을 트리거하도록 추가로 구성된다. 자동화된 어시스턴트의 서버부는 음성 자유 형식 입력을 나타내는 데이터에 기초하여 사용자의 의도를 결정하도록 구성된다. 자동화된 어시스턴트의 서버부는 서드 파티 컴퓨팅 서비스에 의도를 나타내는 데이터를 제공하도록 추가로 구성된다. 의도를 나타내는 데이터는 서드 파티 컴퓨팅 서비스가 해결 정보를 생성하려는 의도를 해결하도록 한다. 자동화된 어시스턴트의 서버부는 또한 의도를 나타내는 데이터의 제공에 응답하여 서드 파티 컴퓨팅 서비스로부터 디스플레이 컨텍스트를 수신하도록 구성된다. 디스플레이 컨텍스트는 인간 대 컴퓨터 대화 세션과 관련된 서드 파티 컴퓨팅 서비스를 위해 유지되고 사용자의 의도 및 해결 정보 중 적어도 하나에 부분적으로 기초하거나 상기 사용자의 의도 및 상기 해결 정보 모두에 기초하여 업데이트된다.

[0025] 여기에 개시된 기술의 이들 및 다른 구현은 선택적으로 다음 특징 중 하나 이상을 포함할 수 있다.

[0026] 일부의 구현 예에서, 그래픽 사용자 인터페이스는 자동화된 어시스턴트의 클라이언트부에 내장 된 웹 브라우저를 포함할 수 있다. 일부의 구현 예에서, 비접촉 상호 작용은 그래픽 사용자 인터페이스의 선택 가능한 요소 작동, 그래픽 사용자 인터페이스에서 렌더링 된 문서의 특정 위치로 스크롤, 및 그래픽 사용자 인터페이스의 일부 확대 중 하나 이상을 포함할 수 있다.

[0027] 일부의 구현 예에서, 디스플레이부를 갖는 클라이언트 장치에서 자동화된 어시스턴트의 클라이언트부를 구현하는 것을 포함하는 하나 이상의 프로세서에 의해 수행되는 방법이 제공된다. 상기 방법은 하나 이상의 서버 컴퓨팅 장치에서 자동화된 어시스턴트의 서버부를 구현하는 단계를 더 포함한다. 사용자는 클라이언트부와 상호 작용하여 사용자, 자동화된 어시스턴트 및 서드 파티 컴퓨팅 서비스 간의 인간 대 컴퓨터 대화 세션에 참여한다. 서드 파티 컴퓨팅 서비스는 구두 대화 상태 머신 및 시각적 대화 상태 머신에 따라 사용자와 연결된다. 상기 방법은 클라이언트부에서, 자동화된 어시스턴트의 서버부를 통해 서드 파티 컴퓨팅 서비스로부터 시각적 대화 상태 머신의 제1 시각적 대화 상태를 나타내는 데이터를 수신하는 단계를 더 포함한다. 상기 방법은 클라이언트부에 의해 제1 시각적 대화 상태에 기초하여 인간 대 컴퓨터 대화 세션과 연관된 그래픽 사용자 인터페이스를 디스플레이부에 렌더링하는 단계를 더 포함한다. 그래픽 사용자 인터페이스는 언어 상태 머신이 제1 시각적 대화 상태에 대응하는 제1 구두 대화 상태에서 제2 구두 대화 상태로 전환되도록 하는 적어도 하나의 그래픽 요소를 포함한다. 또한 상기 방법은 클라이언트부에 의해 사용자에 의한 적어도 하나의 그래픽 요소의 동작을 검출하는 단계를 더 포함한다. 상기 방법은 클라이언트부에 의해 자동화된 어시스턴트의 서버부에 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터를 제공하는 단계를 더 포함한다. 상기 방법은 서버부에 의해 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터를 서드 파티 컴퓨팅 서비스에 제공하는 단계를 더 포함한다. 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터는 서드 파티 컴퓨팅 서비스가 제1 구두 대화 상태에서 제2 구두 대화 상태로 전환되도록 한다.

[0028] 여기에 개시된 기술의 이들 및 다른 구현은 선택적으로 다음 특징 중 하나 이상을 포함할 수 있다.

[0029] 일부의 구현 예에서는 서버부에 의해, 서드 파티 컴퓨팅 서비스로부터, 제2 구두 대화 상태를 나타내는 데이터를 수신하는 단계와, 상기 서버부에 의해, 자동화된 어시스턴트의 클라이언트부에, 제2 구두 대화 상태를 나타내는 데이터를 제공하는 단계를 더 포함할 수 있다. 일부의 구현 예에서, 자동화된 어시스턴트의 서버부에서 서드 파티 컴퓨팅 서비스로의 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터의 제공은 서드 파티 컴퓨팅 서비스가 제1 시각적 대화 상태에서 제2 시각적 대화 상태로 전환되도록 할 수 있다.

[0030] 일부 구현에서, 하나 이상의 프로세서에 의해 적어도 부분적으로 구현된 컴퓨팅 서비스에 의해, 사용자와 자동화된 어시스턴트 간의 인간 대 컴퓨터 대화 세션의 일부로서 자동화된 어시스턴트와 통신하는 컴퓨팅 장치의 사용자의 의도를 나타내는 데이터를 자동화된 어시스턴트로부터 수신하는 단계를 포함하는 방법이 제공된다. 상기 방법은 해결 정보를 생성하려는 사용자의 의도를 해결하는 단계를 더 포함한다. 상기 방법은 인간 대 컴퓨터 대화 세션과 관련된 컴퓨팅 서비스를 위해 유지되는 디스플레이 컨텍스트를 업데이트하는 단계를 더 포함한다. 상기 업데이트는 의도 및 해결 정보 중 적어도 하나에 부분적으로 기초하거나 상기 사용자의 의도 및 상기 해결 정보 모두에 기초한다. 상기 방법은 디스플레이 컨텍스트를 나타내는 데이터를 자동화된 어시스턴트에게 제공하는 단계를 더 포함한다. 상기 디스플레이 컨텍스트를 나타내는 데이터는 자동화된 어시스턴트에 의해 컴퓨팅 장치에 제공되고 컴퓨팅 장치에서 실행되는 어시스턴트 애플리케이션이 사용자와 어시스턴트 애플리케이션의 그래픽 사용자 인터페이스 간의 비접촉 상호 작용을 트리거하게 한다.

- [0031] 여기에 개시된 기술의 이들 및 다른 구현은 선택적으로 다음 특징 중 하나 이상을 포함할 수 있다.
- [0032] 일부의 구현 예에서, 그래픽 사용자 인터페이스는 어시스턴트 애플리케이션에 내장 된 웹 브라우저를 포함할 수 있다. 일부의 구현 예에서, 비접촉 상호 작용은 그래픽 사용자 인터페이스의 선택 가능한 요소의 작동, 그래픽 사용자 인터페이스에서 렌더링된 문서의 특정 위치로 스크롤, 그래픽 사용자 인터페이스의 일부 확대 중 하나 이상을 포함할 수 있다. 이들 구현 중 일부에서, 사용자의 의도를 나타내는 데이터는 컴퓨팅 장치에서 사용자에게 의해 제공되는 음성 자유 형식 입력의 음성 인식 출력을 포함할 수 있다. 일부의 구현 예에서, 방법은 음성 인식 출력에 기초하여 사용자의 의도를 컴퓨팅 서비스에 의해 결정하는 단계를 더 포함할 수 있다.
- [0033] 다른 구현은 하나 이상의 프로세서(예를 들어, 중앙 처리 장치(들) (CPU(들)), 그래픽 처리 장치(들)(GPU), 및/또는 텐서 프로세싱 유닛(들) (TPU(s)))은 상기 및/또는 본 명세서의 다른 곳에서 설명 된 방법들 중 하나 이상과 같은 방법을 수행한다. 또 다른 구현은 상기 및/또는 본 명세서의 다른 곳에서 설명 된 하나 이상의 방법과 같은 방법을 수행하기 위해 저장된 명령을 실행하도록 동작 가능한 하나 이상의 프로세서를 포함하는 하나 이상의 컴퓨터의 시스템을 포함할 수 있다.
- [0034] 전술한 개념 및 본 명세서에 더 상세히 설명된 추가 개념의 모든 조합은 본 명세서에 개시된 주제의 일부인 것으로 고려된다는 것을 이해해야한다. 예를 들어, 본 개시의 끝에 나타나는 청구된 주제의 모든 조합은 여기에 개시된 주제의 일부인 것으로 고려될 수 있다.

**도면의 간단한 설명**

- [0035] 도 1은 본 명세서에 개시된 실시 예가 구현될 수 있는 예시적인 환경의 블록도이다.
- 도 2는 다양한 구현에 따라 자동화된 어시스턴트의 클라이언트부에서 구현될 수 있는 컴포넌트의 예를 도시한다.
- 도 3 및 도 4는 본 명세서에 설명된 기술이 특정 시나리오에서 적용되는 예를 도시한다.
- 도 5 및 도 6은 본 명세서에 설명된 기술이 다른 시나리오에서 적용되는 예를 도시한다.
- 도 7은 본 명세서에 설명된 기술이 또 다른 시나리오에서 적용되는 예를 도시한다.
- 도 8 및 도 9는 본 명세서에 개시된 실시 예에 따른 예시적인 방법을 나타내는 흐름도를 도시한다.
- 도 10은 컴퓨팅 장치의 예시적인 아키텍처를 도시한다.

**발명을 실시하기 위한 구체적인 내용**

- [0036] 이제 도 1을 참조하면, 여기에 개시된 기술이 구현될 수 있는 예시적인 환경이 도시된다. 예시적인 환경은 하나 이상의 클라이언트 컴퓨팅 장치(106)를 포함한다. 각 클라이언트 장치(106)는 자동화된 어시스턴트의 "클라이언트부"로도 지칭 될 수 있는 자동화된 어시스턴트 클라이언트(108)의 각각의 인스턴스(instance)를 실행할 수 있다. 여기서, 총칭하여 자동화된 어시스턴트의 "서버부"라고도 하는 하나 이상의 클라우드 기반 자동 어시스턴트 컴포넌트(component)(119)는, 하나 이상의 컴퓨팅 시스템(총칭하여 "클라우드"컴퓨팅 시스템으로 지칭됨)에서 구현되고, 일반적으로 '114'로 표시된 하나 이상의 로컬 및/또는 광역 네트워크 (예를 들어, 인터넷)를 통해 클라이언트 장치(106)와 통신으로 결합된다.
- [0037] 다양한 구현에서, 자동화된 어시스턴트 클라이언트 (108)의 인스턴스는 하나 이상의 클라우드 기반 자동 어시스턴트 컴포넌트(119)와의 상호 작용을 통해, 사용자의 관점에서, 사용자가 인간 대 컴퓨터 대화에 참여할 수 있는 자동화된 어시스턴트(120)의 논리적 인스턴스를 형성할 수 있다.
- [0038] 그러한 자동화된 어시스턴트(120)의 한 예가, 도 1에 점선으로 도시되어 있다. 따라서 클라이언트 장치(106)에서 실행되는 자동화된 어시스턴트 클라이언트(108)와 소통하는 각 사용자는, 사실상 자신의 자동화된 어시스턴트(120)의 논리적 인스턴스를 상대하고 있음을 이해해야 한다.
- [0039] 간결함과 단순함을 위해, 본 명세서에서 특정 사용자를 "서비스하는"것으로 사용되는 자동 어시스턴트라는 용어는 사용자에게 의해 작동되는 클라이언트 장치(106)에서 실행되는 자동 어시스턴트 클라이언트(108)와 하나 이상의 클라우드 기반 자동 어시스턴트 컴포넌트(119)의 조합을 지칭할 것이다. 또한 일부의 구현 예에서, 자동화된 어시스턴트(120)는 사용자가 자동화된 어시스턴트(120)의 특정 인스턴스에 의해 실제로 "서비스"를 받는지 여부에 관계없이 임의의 사용자로부터의 요청에 응답할 수 있다는 것임을 이해해야 한다.

- [0040] 하나 이상의 클라이언트 장치(106), 예를 들어 데스크톱 컴퓨팅 장치, 랩톱 컴퓨팅 장치, 태블릿 컴퓨팅 장치, 모바일 전화 컴퓨팅 장치, 사용자 차량의 컴퓨팅 장치(예: 차량용 통신 시스템, 차량용 엔터테인먼트 시스템, 차량용 내비게이션 시스템), 독립형 대화형 스피커(일부 경우 비전(vision) 센서를 포함할 수 있음), 스마트 TV(또는 자동화된 어시스턴트 기능이 있는 네트워크 동글이 장착된 표준 텔레비전) 및/또는 컴퓨팅 장치를 포함하는 사용자의 웨어러블 장치(예: 컴퓨팅 장치가 있는 사용자의 시계, 컴퓨팅 장치가 있는 사용자의 안경, 가상 또는 증강 현실 컴퓨팅 장치)를 포함할 수 있다.
- [0041] 추가 및/또는 대체되는 클라이언트 컴퓨팅 장치가 제공될 수 있다. 독립형 대화형 스피커(또는 "스마트 스피커")와 같은 일부 클라이언트 장치(106)는 주로 사용자와 자동화된 어시스턴트(120) 간의 대화를 용이하게 하도록 설계된 어시스턴트 장치의 형태를 취할 수 있다.
- [0042] 일부의 구현 예에서, 클라이언트 장치(106)는 하나 이상의 시야를 갖는 하나 이상의 비전 센서(107)를 구비할 수 있지만, 이것이 필수는 아니다. 비전 센서(들)(107)는 디지털 카메라, 수동 적외선 ("PIR") 센서, 입체 카메라, RGBd 카메라와 같이 다양하게 형성될 수 있다. 예를 들어 하나 이상의 비전 센서(107)는 이미지 캡처 모듈(111)에 사용되어 클라이언트 장치(106)가 배치된 환경의 이미지 프레임(정지 이미지 또는 비디오)을 캡처하기 위해 사용될 수 있다. 그 다음, 이러한 이미지 프레임은 이미지 프레임에 포함된 사용자 제공 시각적 큐(User-provided visual cue)를 검출하기 위해, 예를 들어 시각적 큐(visual cue)모듈(112<sub>i</sub>)에 의해 분석될 수 있다. 이러한 시각적 큐는 손 제스처, 특정 기준점을 향한 시선, 얼굴 표정, 사용자에게 의한 미리 정의된 움직임 등을 포함할 수 있지만 이에 제한되지 않는다. 이러한 감지된 시각적 큐는 자동화된 어시스턴트(120) 호출 및/또는 자동화된 어시스턴트(120)가 다양한 동작들을 유발하게 하는 등 다양한 목적으로 사용될 수 있다.
- [0043] 본 명세서에서 더 상세히 설명되는 바와 같이, 자동화된 어시스턴트(120)는 하나 이상의 클라이언트 장치(106)의 사용자 인터페이스 입력 및 출력 장치를 통해 하나 이상의 사용자와 인간 대 컴퓨터 대화 세션에 참여한다. 일부의 구현 예에서, 자동화된 어시스턴트(120)는 클라이언트 장치(106) 중 하나의 하나 이상의 사용자 인터페이스 입력 장치를 통해 사용자에게 의해 제공되는 사용자 인터페이스 입력에 응답하여 사용자와 인간 대 컴퓨터 대화 세션에 참여할 수 있다. 이러한 구현에서, 사용자 인터페이스 입력은 명시적으로 자동화된 어시스턴트(120)로 연결된다. 예를 들어, 사용자는 자동화된 어시스턴트(120)가 타이핑된 텍스트를 능동적으로 듣거나 모니터링하도록 하기 위해 "OK, Assistant" 또는 "Hey, Assistant"와 같은 미리 결정된 호출 문구를 구두로 제공(예를 들어, 타이핑, 말하기)할 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, 자동화된 어시스턴트(120)는 하나 이상의 검출된 시각적 큐에 기초하여 단독으로 또는 구두 호출 문구와 조합하여 호출될 수 있다.
- [0044] 일부의 구현 예에서, 자동화된 어시스턴트(120)는 사용자 인터페이스 입력이 자동화된 어시스턴트(120)에 명시적으로 지시되지 않은 경우에도 사용자 인터페이스 입력에 응답하여 인간 대 컴퓨터 대화 세션에 참여할 수 있다. 예를 들어, 자동화된 어시스턴트(120)는 사용자 인터페이스 입력의 내용을 검사하고 사용자 인터페이스 입력에 존재하는 특정 용어에 응답 및/또는 다른 단서에 기초하여 대화 세션에 참여할 수 있다. 다수의 구현 예로서, 자동화된 어시스턴트(120)는 음성 인식을 활용하여 사용자의 발화를 텍스트로 변환하고, 그 텍스트에 따른 반응(예를 들어 검색 결과나 일반 정보의 제공)하거나, 하나 이상의 응답 동작(예를 들어 음악 재생, 게임 실행, 음식 주문 등)을 수행할 수 있다. 일부의 구현 예에서, 자동화된 어시스턴트(120)는 발화를 텍스트로 변환하지 않고 발화에 추가적으로 또는 대안적으로 응답할 수 있다. 예를 들어, 자동화된 어시스턴트(120)는 음성 입력을 임베딩, 엔티티(entity) 표현(음성 입력에 존재하는 엔티티/엔티티들을 나타냄) 및/또는 다른 "비텍스트" 표현으로 변환할 수 있으며, 그러한 비텍스트 표현으로 동작할 수 있다. 따라서 음성 입력으로부터 변환된 텍스트에 기초하여 동작하는 것으로 본 명세서에 설명된 구현은 추가적으로 및/또는 대안적으로 음성 입력에 대해 직접 및/또는 음성 입력의 다른 비텍스트 표현에 대해 동작할 수 있다.
- [0045] 클라우드 기반 자동 어시스턴트 컴포넌트(119)를 운영하는 클라이언트 컴퓨팅 장치(106) 및 컴퓨팅 장치(들) 각각은, 데이터 및 소프트웨어 애플리케이션의 저장을 위한 하나 이상의 메모리, 데이터에 액세스하고 애플리케이션을 실행하기 위한 하나 이상의 프로세서, 및 네트워크를 통한 통신을 용이하게 하는 다른 구성 요소를 포함할 수 있다. 클라이언트 컴퓨팅 장치(106) 및/또는 자동화된 어시스턴트(120)에 의해 수행되는 동작은 다수의 컴퓨터 시스템에 분산될 수 있다. 자동화된 어시스턴트(120)는 예를 들어 네트워크를 통해 서로 연결된 하나 이상의 위치에 있는 하나 이상의 컴퓨터에서 실행되는 컴퓨터 프로그램으로서 구현될 수 있다.
- [0046] 진술한 바와 같이, 다양한 구현에서, 클라이언트 장치(106)는 자동화된 어시스턴트 클라이언트(108) 또는 자동화된 어시스턴트(120)의 "클라이언트부"를 동작시킬 수 있다. 다양한 구현에서, 자동화된 어시스턴트 클라이언트



트(108)는 음성 캡처 모듈(110), 전술한 이미지 캡처 모듈(111), 시각적 큐 모듈(112<sub>1</sub>) 및/또는 호출 모듈(113)을 포함할 수 있다. 다른 구현에서, 음성 캡처 모듈(110), 이미지 캡처 모듈(111), 시각적 큐 모듈(112<sub>1</sub>), 및/또는 호출 모듈(113)의 하나 이상의 측면은 자동화된 어시스턴트 클라이언트 (108)와 별도로, 예를 들어 하나 이상의 클라우드 기반 자동 어시스턴트 컴포넌트(119)에 의해 구현 될 수 있다. 예를 들어, 도 1에는 이미지 데이터에서 시각적 큐를 검출 할 수 있는 클라우드 기반 시각적 큐 모듈(112<sub>2</sub>)도 있다.

[0047] 다양한 구현 예에서, 하드웨어 및 소프트웨어의 임의의 조합을 사용하여 구현될 수 있는 음성 캡처 모듈(110)은 사용자의 발화(들)의 오디오 녹음을 캡처하기 위해, 마이크로폰(109) 또는 다른 압력 센서와 같은 하드웨어와 인터페이스를 수행할 수 있다. 다양한 목적으로 이 오디오 녹음에 대한 다양한 유형의 처리가 수행될 수 있다. 일부의 구현 예에서, 하드웨어 또는 소프트웨어의 임의의 조합을 사용하여 구현될 수 있는 이미지 캡처 모듈(111)은 카메라(107)와 인터페이스하여 비전 센서(107)의 시야에 해당하는 하나 이상의 이미지 프레임(예 : 디지털 사진)을 캡처하도록 구성될 수 있다.

[0048] 다양한 구현에서, 시각적 큐 모듈(112<sub>1</sub>)(및/또는 클라우드 기반 시각적 큐 모듈 112<sub>2</sub>)은 하드웨어 또는 소프트웨어의 임의의 조합을 사용하여 구현 될 수 있으며, 이미지 캡처 모듈(111)에 의해 제공되는 하나 이상의 이미지 프레임을 분석하여 하나 이상의 이미지 프레임에 걸쳐서, 또는 상기 하나 이상의 이미지 프레임 내에서 캡처된 하나 이상의 시각적 큐를 검출하도록 구성 될 수 있다. 시각적 큐 모듈(112<sub>1</sub>)은 시각적 큐를 검출하기 위해 다양한 기술을 사용할 수 있다. 예를 들어, 도 1에서, 시각적 큐 모듈(112<sub>1</sub>)은 시각적 큐 모델 데이터베이스(114) (예를 들어 클라우드와 같이, 클라이언트 장치(106)와 통합 및/또는 클라이언트 장치(106)로부터 원격으로 호스팅 될 수 있음)와 통신으로 결합된다. 시각적 큐 모델 데이터베이스(114)는, 예를 들어 이미지 프레임에서 검출된 사용자 제공 시각적 큐를 나타내는 출력이 생성되도록 훈련된 하나 이상의 인공 지능(또는 기계 학습) 모델을 포함할 수 있다.

[0049] 음성 캡처 모듈(110)은 앞서 언급 한 바와 같이, 예를 들어 마이크로폰(109)을 통해 사용자의 음성을 캡처하도록 구성 될 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, 음성 캡처 모듈(110)은 캡처된 오디오를 텍스트 및/또는 다른 표현 또는 임베딩(embeddings)으로, 예를 들어 음성-텍스트("STT") 처리 기술을 사용하여 변환하도록 추가로 구성될 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, 음성 캡처 모듈(110)은 예를 들어 하나 이상의 음성 합성기를 사용하여 텍스트를 컴퓨터 합성 음성으로 변환하도록 구성될 수 있다. 그러나 클라이언트 장치(106)는 컴퓨팅 리소스(예를 들어, 프로세서 사이클, 메모리, 배터리 등) 측면에서 상대적으로 제한될 수 있기 때문에, 클라이언트 장치(106)에 형성된 음성 캡처 모듈(110)은 한정된 수의 상이한 음성 문구, 특히 자동화된 어시스턴트(120)를 호출하는 문구를 텍스트로(또는 저차원 임베딩과 같은 다른 형태로) 변환하도록 구성될 수 있다. 다른 음성 입력은 클라우드 기반 텍스트 음성 변환("TTS") 모듈(116) 및/또는 클라우드 기반 STT 모듈(117)을 포함할 수 있는 클라우드 기반 자동 어시스턴트 컴포넌트(119)로 전송 될 수 있다.

[0050] 다양한 구현에서, 호출 모듈(113)은, 예를 들어 음성 캡처 모듈(110) 및/또는 시각적 큐 모듈(112<sub>1</sub>)에 의해 제공되는 출력에 기초하여 자동화된 어시스턴트(120)를 호출할지 여부를 결정하도록 구성될 수 있다(일부의 구현 예에서 단일 모듈 내에서 이미지 캡처 모듈(111)과 결합될 수 있음). 예를 들어, 호출 모듈(113)은 사용자의 발화가, 자동화된 어시스턴트(120)와 인간 대 컴퓨터 대화 세션을 시작해야 하는 호출 문구로서 자격이 있는지 여부를 결정할 수 있다. 일부의 구현 예에서, 호출 모듈(113)은 시각적 큐 모듈(112<sub>1</sub>)에 의해 검출된 하나 이상의 시각적 큐와 함께 오디오 녹음 또는 오디오 녹음으로부터 추출된 특징의 벡터(예를 들어, 임베딩)와 같은 사용자의 발화를 나타내는 데이터를 분석할 수 있다. 일부의 구현 예에서, 특정 시각적 큐가 또한 검출될 때, 음성 발화에 응답하여 자동화된 어시스턴트(120)를 호출할지 여부를 결정하기 위해 호출 모듈(113)에 의해 사용되는 임계값이 낮아질 수 있다. 결과적으로, 사용자가 적절한 호출 문구 "OK assistant"와는 다르지만 음성학적으로 다소 유사한 음성 발화를 제공하더라도 시각적 큐와 함께 감지될 때 발화가 적절한 호출로 받아들여질 수 있다 (예: 화자가 손을 흔드는 경우, 화자가 비전 센서(107)를 직접 응시하는 등).

[0051] 일부의 구현 예에서, 온-디바이스 호출 모델은 발화 및/또는 시각적 큐(들)가 호출로 적격인지 여부를 결정하기 위해 호출 모듈(113)에 의해 사용될 수 있다. 이러한 온-디바이스 호출 모델은 호출 문구/제스처의 변형을 감지하도록 훈련 될 수 있다. 예를 들어, 일부의 구현 예에서, 온-디바이스 호출 모델(예를 들어 하나 이상의 신경망)은 사용자의 발화와 동시에 상기 발화에 대한 오디오 녹음(또는 추출된 특징 벡터) 뿐만 아니라 캡처 된 하나 이상의 이미지 프레임 및/또는 감지 된 시각적 큐를 나타내는 데이터를 각각 포함하는 훈련 예제를 사용하여

훈련될 수 있다.

- [0052] 클라우드 기반 텍스트 음성 변환(TTS) 모듈(116)(이하 TTS 모듈)은 텍스트 데이터(예를 들어, 자동화된 어시스턴트(120)에 의해 공식화 된 자연어 응답)를 컴퓨터 생성 음성 출력으로 변환하기 위해 클라우드의 사실상 무한한 자원을 활용하도록 구성될 수 있다. 일부의 구현 예에서, 상기 TTS 모듈(116)은 예를 들어 하나 이상의 스피커를 사용하여 직접 출력되도록 컴퓨터 생성 음성 출력을 클라이언트 장치(106)에 제공 할 수 있다. 다른 구현들에서, 자동화된 어시스턴트(120)에 의해 생성된 텍스트 데이터(예를 들어, 자연어 응답)는 음성 캡처 모듈(110)에 제공될 수 있으며, 이는 텍스트 데이터를 로컬로 출력되는 컴퓨터 생성 음성으로 변환될 수 있다.
- [0053] 클라우드 기반 음성 텍스트 변환(STT) 모듈(117)(이하 STT 모듈)은 클라우드의 사실상 무한한 리소스를 활용하여 음성 캡처 모듈(110)에 의해 캡처된 오디오 데이터를 텍스트로 변환하도록 구성될 수 있으며, 이는 자연어 이해 모듈(135)에 제공될 수 있다. 일부의 구현 예에서, 클라우드 기반 상기 STT 모듈(117)은 음성의 오디오 녹음을 하나 이상의 음소로 변환한 다음, 하나 이상의 음소를 텍스트로 변환할 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, STT 모듈(117)은 상태 디코딩 그래프를 사용할 수 있다. 일부의 구현 예에서, STT 모듈(117)은 사용자의 발화에 대한 복수의 후보 텍스트 해석을 생성할 수 있다. 일부의 구현 예에서, STT 모듈(117)은 동시적으로 검출된 시각적 큐가 있는지 여부에 따라 특정 후보 텍스트 해석에 더 높은 가중치를 부여하거나 편중(bias)될 수 있다. 예를 들어 두 후보 텍스트 해석의 신뢰도 점수가 비슷하다고 가정한다. 종래의 자동화된 어시스턴트(120)에 의해, 사용자는 이러한 후보 텍스트 문장 사이를 명확하게 하도록 요청받을 수 있다. 그러나 본 개시의 선택된 양상들로 구성된 자동화된 어시스턴트(120)에서, 하나 이상의 검출 된 시각적 큐가 "동점 깨기(break the tie)"를 위해 사용될 수 있다.
- [0054] 자동화된 어시스턴트(120)(특히, 클라우드 기반 자동화된 어시스턴트 컴포넌트(119))는 의도 매칭기(135), 전술한 TTS 모듈(116), 전술한 STT 모듈(117) 및 아래에서 더 상세히 설명되는 다른 컴포넌트를 포함할 수 있다. 일부의 구현 예에서, 자동화된 어시스턴트(120)의 모듈 및/또는 모듈 중 하나 이상이 생략, 결합 및/또는 자동화된 어시스턴트(120)와 별개인 컴포넌트에서 구현 될 수 있다. 일부의 구현 예에서, 프라이버시를 보호하기 위해, 자연어 프로세서(122), TTS 모듈(116), STT 모듈(117) 등과 같은 자동화된 어시스턴트(120)의 구성 요소 중 하나 이상은 클라이언트 장치(106)에 구현될 수 있다(예를 들어, 클라우드를 배제하기 위하여).
- [0055] 일부의 구현 예에서, 자동화된 어시스턴트(120)는 자동화된 어시스턴트(120)와의 인간 대 컴퓨터 대화 세션 동안 클라이언트 장치(106) 중 하나의 사용자에게 의해 생성된 다양한 입력에 응답하여 응답 콘텐츠를 생성한다. 자동화된 어시스턴트(120)는 대화 세션의 일부로서 사용자에게 프리젠테이션을 위해(예를 들어, 사용자의 클라이언트 장치로부터 분리될 때 하나 이상의 네트워크를 통해) 응답 콘텐츠를 제공할 수 있다. 예를 들어, 자동화된 어시스턴트(120)는 클라이언트 장치(106)를 통해 제공되는 자유 형식 자연어 입력에 응답하여 응답 콘텐츠를 생성할 수 있다. 본 명세서에서 사용되는 바와 같이, 자유 형식 입력은 사용자에게 의해 공식화되고 사용자가 선택하기 위해 제시된 옵션 그룹에 제한되지 않는 입력이다.
- [0056] 본 명세서에서 사용되는 바와 같이, 대화 세션은 논리적으로 자체적으로 포함된 사용자와 자동화된 어시스턴트(120) 사이의 하나 이상의 메시지들을 포함할 수 있다. 자동화된 어시스턴트(120)는 세션 간의 시간 경과, 세션 간의 사용자 컨텍스트(context) 변경, 사용자와 자동화된 어시스턴트 간의 대화를 제외하고, 사용자와 클라이언트 장치 간에 개입하는 하나 이상의 상호 작용 감지, 세션 간 클라이언트 장치 잠금/슬립, 자동화된 어시스턴트(120)의 하나 이상의 인스턴스와 인터페이스하는 데 사용되는 클라이언트 장치의 변경 등등, 다양한 신호에 기초하여 사용자와의 다중 대화 세션을 구별할 수 있다.
- [0057] 의도 매칭기(135)는 사용자에게 의해 제공되는 입력(들)(예를 들어, 음성 발화, 시각적 큐 등) 및/또는 센서 신호, 온라인 신호(예: 웹 서비스에서 얻은 데이터) 등등에 기초하여 사용자의 의도를 결정하도록 구성될 수 있다. 일부의 구현 예에서, 의도 매칭기(135)는 자연어 프로세서(122) 및 전술한 클라우드 기반 시각적 큐 모듈(112<sub>2</sub>)을 포함할 수 있다. 다양한 구현에서, 클라우드 기반 시각적 큐 모듈(112<sub>2</sub>)은, 클라우드 기반 시각적 큐 모듈(112<sub>2</sub>)이 더 많은 리소스를 가질 수 있다는 점을 제외하고는 시각적 큐 모듈(112<sub>1</sub>)과 유사하게 동작 할 수 있다. 특히, 클라우드 기반 시각적 큐 모듈(112<sub>2</sub>)은 사용자의 의도를 결정하기 위해, 의도 매칭기(135)에 의해 단독으로 또는 다른 신호와 결합하여 사용될 수 있는 시각적 큐를 검출 할 수 있다.
- [0058] 자연어 프로세서(122)는, 클라이언트 장치(106)를 통해 사용자(들)에 의해 생성된 자연어 입력을 처리하도록 구성될 수 있고 자동화된 어시스턴트(120)의 하나 이상의 다른 구성 요소에 의해 사용하기 위해 주석이 달린 출력(예를 들어, 텍스트 형태)을 생성할 수 있다. 예를 들어, 자연어 프로세서(122)는 클라이언트 장치(106)의 하나

이상의 사용자 인터페이스 입력 장치를 통해 사용자에게 의해 생성되는 자연어 자유 형식 입력을 처리할 수 있다. 생성된 주석이 달린 출력은, 자연어 입력의 하나 이상의 주석과 자연어 입력의 용어 중 하나 이상(예를 들어, 모두)을 포함한다.

- [0059] 일부의 구현 예에서, 자연어 프로세서(122)는 자연어 입력에서 다양한 유형의 문법 정보를 식별하고 주석을 달도록 구성된다. 예를 들어, 자연어 프로세서(122)는 개별 단어를 형태소로 분리하고 /하거나 형태소, 예를 들어 그 클래스에 주석을 달 수 있는 형태소 모듈을 포함할 수 있다. 자연어 프로세서(122)는 또한 그들의 문법적 역할과 함께 용어에 주석을 달도록 구성된 음성 태거의 일부를 포함할 수 있다. 예를 들어, 품사 태거는 각 용어에 "명사", "동사", "형용사", "대명사" 등과 같은 품사로 태그를 지정할 수 있다. 또한, 예를 들어, 일부의 구현 예에서 자연어 프로세서(122)는 자연어 입력에서 용어들 간의 구문 관계를 결정하도록 구성된 종속성 파서(parser)(도시되지 않음)를 추가적으로 및/또는 대안적으로 포함할 수 있다. 예를 들어, 종속성 파서는 어떤 용어가, 다른 용어, 주제 및 문장의 동사 등(예 : 구문 분석 트리)을 수정하는지 결정할 수 있으며, 이러한 종속성에 대한 주석을 만들 수 있다.
- [0060] 일부의 구현 예에서, 자연어 프로세서(122)는 추가적으로 및/또는 대안적으로 사람(예 : 문학 인물, 유명인, 유명인 등 포함), 조직, 위치(실제 및 가상) 등에 대한 참조와 같은 하나 이상의 세그먼트에서 엔티티 참조에 주석을 달도록 구성되는 엔티티 태거(tagger)(도시되지 않음)를 포함할 수 있다. 일부의 구현 예에서, 엔티티에 대한 데이터는 지식 그래프(도시되지 않음)와 같은 하나 이상의 데이터베이스에 저장될 수 있다. 일부의 구현 예에서, 지식 그래프는 노드를 연결하고 엔티티 간의 관계를 나타내는 에지(edge) 뿐만 아니라, 알려진 엔티티(및 일부 경우에 엔티티 속성)를 나타내는 노드를 포함할 수 있다. 예를 들어, "바나나"노드는 "과일"노드에 연결될 수 있으며(예를 들어, 차일드(child)로서), 차례로 연결되어(예를 들어, 차일드로서) "생산"및/또는 "음식"노드에 연결될 수 있다. 또 다른 예로, "Hypothetical Cafe"라는 레스토랑은 주소, 제공되는 음식 유형, 시간, 연락처 정보 등과 같은 속성도 포함하는 노드로 표현될 수 있다. "Hypothetical Cafe"노드는 일부의 구현 예에서 에지(예: 자식-부모 관계를 나타내는)에 의해 하나 이상의 다른 노드, 예를 들어 "레스토랑"노드, "비즈니스"노드, 상기 레스토랑이 위치한 도시 및/또는 주를 나타내는 노드 등에 연결될 수 있다.
- [0061] 자연어 프로세서(122)의 엔티티 태거는 높은 수준의 입도(예 : 사람과 같은 엔티티 클래스에 대한 모든 참조를 식별할 수 있도록 하기 위해) 및/또는 더 낮은 수준의 입도(예 : 특정 사람과 같은 특정 엔티티에 대한 모든 참조를 식별할 수 있도록 하기 위해)에서 개체에 대한 참조에 주석을 달 수 있다. 엔티티 태거는 특정 엔티티를 해결하기 위해 자연어 입력의 내용에 의존할 수 있다. 그리고/또는 특정 엔티티를 해결하기 위해 지식 그래프 또는 다른 엔티티 데이터베이스와 선택적으로 통신할 수 있다.
- [0062] 일부의 구현 예에서, 자연어 프로세서(122)는 추가적으로 및/또는 대안적으로 하나 이상의 컨텍스트 단서에 기초하여 동일한 엔티티에 대한 참조를 그룹화하거나 "클러스터"하도록 구성된 상호 참조 해석기(도시되지 않음)를 포함할 수 있다. 예를 들어, 상호 참조 해석기는 자연어 입력 "지난번에 그곳에서 먹었을 때 나는 Hypothetical Cafe가 좋았습니다"에서 "그곳"이라는 용어를 "Hypothetical Cafe"로 해석하는 데 활용될 수 있다.
- [0063] 일부의 구현 예에서, 자연어 프로세서(122)의 하나 이상의 컴포넌트는 자연어 프로세서(122)의 하나 이상의 다른 컴포넌트로부터의 주석에 의존할 수 있다. 예를 들어, 일부의 구현 예에서 명명된 엔티티 태거는 특정 엔티티에 대한 모든 언급들에 주석을 달 때 상호 참조 해석기 및/또는 종속성 파서의 주석에 의존할 수 있다. 또한, 예를 들어 일부의 구현 예에서 상호 참조 해석기는 동일한 엔티티에 대한 참조를 클러스터링할 때 종속성 파서의 주석에 의존할 수 있다. 일부의 구현 예에서, 특정 자연어 입력을 처리할 때, 자연어 프로세서(122)의 하나 이상의 구성 요소는 관련된 이전 입력 및/또는 하나 이상의 주석을 결정하기 위한 특정 자연어 입력이 아닌 관련된 다른 데이터를 사용할 수 있다.
- [0064] 의도 매칭기(135)는 사용자의 의도를 결정하기 위해 다양한 기술을 사용할 수 있다. 예를 들어 자연어 프로세서(122)로부터의 출력(자연어 입력의 주석 및 용어를 포함할 수 있음)에 기초할 수 있다. 그리고/또는 시각적 큐 모듈(예 : 112<sub>1</sub> 및/또는 112<sub>2</sub>)로부터의 출력에 기초할 수 있다. 일부의 구현 예에서, 의도 매칭기(135)는, 예를 들어 문법, 시각적 큐 및 응답 동작(또는 보다 일반적으로 의도들) 사이의 복수의 매핑을 포함하는 하나 이상의 데이터베이스(도시되지 않음)에 액세스할 수 있다. 많은 경우에 이러한 문법은 시간이 지남에 따라 선택 및/또는 학습될 수 있으며 사용자의 가장 일반적인 의도를 나타낼 수 있다. 예를 들어, 하나의 문법, "play <artist>"는 <artist> 에 의한 음악이 사용자에게 의해 작동되는 클라이언트 장치(106)에서 재생되게 하는 반응적 동작을 호출하는 의도에 매핑될 수 있다. 또 다른 문법 인 "[날씨 예보] 오늘"은 "오늘 날씨는 어때?", "오늘의



예보는 어때?"와 같은 사용자 질의와 일치할 수 있다.

- [0065] 문법에 추가하여 또는 대신에, 일부의 구현 예에서, 의도 매칭기(135)는 단독으로 또는 하나 이상의 문법 및/또는 시각적 큐와 조합하여 하나 이상의 훈련된 기계 학습 모델을 사용할 수 있다.
- [0066] 이러한 훈련된 기계 학습 모델은 또한 하나 이상의 데이터베이스에 저장될 수 있으며, 예를 들어 사용자의 발화 및/또는 감지된 사용자 제공 시각적 큐를 나타내는 데이터를 축소된 차원 공간에 삽입하여 의도를 식별하도록 훈련될 수 있다. 예를 들어 유클리드 거리, 코사인 유사성 등과 같은 기술을 사용하여 어떤 다른 임베딩(그로 인한 의도)이 가장 근접한지 결정한다.
- [0067] "play <artist>"예제 문법에서 볼 수 있듯이 일부 문법에는 슬롯 값(또는 "매개 변수")으로 채울 수 있는 슬롯(예 : <artist>)이 있다. 슬롯 값은 다양한 방식으로 결정될 수 있다. 종종 사용자는 사전에 슬롯 값을 제공한다. 예를 들어 "Order me a <topping> pizza"라는 문법의 경우 사용자는 "order me a sausage pizza"라는 문구를 말할 수 있다. 이 경우 슬롯 중지 '>'가 자동으로 채워질 수 있다. 추가적으로 또는 대안적으로, 사용자가 사전에 슬롯 값을 제공하지 않고, 슬롯 값으로 채워질 슬롯을 포함하는 문법을 호출하는 경우, 자동화된 어시스턴트(120)는 사용자로부터 이러한 슬롯 값을 요청할 수 있다(예를 들어, "어떤 유형의 크리스트 피자를 원하십니까?"). 일부의 구현 예에서, 슬롯은 시각적 큐 모듈(112<sub>1-2</sub>)에 의해 검출된 시각적 큐에 기초한 슬롯 값으로 채워질 수 있다. 예를 들어, 사용자는 클라이언트 장치(106)의 시각 센서(107)에 손가락 세 개를 들고 "고양이 그릇 주문"과 같은 말을 할 수 있다. 또는 사용자가 특정 영화의 DVD 케이스를 들고 "이와 같은 영화 더 찾아줘"와 같은 말을 할 수 있다.
- [0068] 일부의 구현 예에서, 자동화된 어시스턴트(120)는 사용자와 하나 이상의 서드 파티 컴퓨팅 서비스(140)(또는 "서드 파티 에이전트"또는 "에이전트") 사이의 중개자 역할을 수행할 수 있다. 이러한 서드 파티 컴퓨팅 서비스(140)는 입력을 수신하고 응답 출력을 제공하는 독립적인 소프트웨어 프로세스일 수 있다. 일부 서드 파티 컴퓨팅 서비스는 예를 들어 클라우드 기반 자동 어시스턴트 컴포넌트(119)를 작동하는 것과 별개인 컴퓨팅 시스템에서 작동하거나 작동하지 않을 수 있는 서드 파티 애플리케이션의 형태를 취할 수 있다. 의도 매칭기(135)에 의해 식별될 수 있는 한 종류의 사용자 의도는 서드 파티 컴퓨팅 서비스(140)와 참여하는 것이다. 예를 들어, 자동화된 어시스턴트(120)는 스마트 장치를 제어하기 위한 서비스에 대한 애플리케이션 프로그래밍 인터페이스("API")에 대한 액세스를 제공할 수 있다. 사용자는 자동화된 어시스턴트(120)를 호출하고 "난방을 켜고 싶다"와 같은 명령을 제공할 수 있다. 의도 매칭기(135)는 이 명령을 자동화된 어시스턴트(120)가 서드 파티 서비스에 참여하도록 트리거하는 문법에 매핑하여 사용자의 난방이 켜지도록 할 수 있다. 서드 파티 서비스는 가열을 켜는 명령을 이행(또는 "해결")하기 위해 채워져야 하는 슬롯의 최소 목록을 자동화된 어시스턴트(120)에 제공할 수 있다. 이 예에서, 슬롯은 가열이 설정될 온도 및 가열이 켜질 기간을 포함할 수 있다. 자동화된 어시스턴트(120)는 슬롯에 대한 파라미터를 요청하는 자연어 출력을 생성하고 사용자에게(클라이언트 장치(106)를 통해) 제공할 수 있다.
- [0069] 이행 모듈(124)은 의도 매칭기(135)에 의해 출력되는 예측/추정 의도 뿐만 아니라 연관된 슬롯 값(사용자에 의해 사전에 제공되거나 사용자로부터 요청됨)을 수신하고 그 의도를 이행(또는 "해결")하도록 구성될 수 있다. 다양한 구현에서, 사용자 의도의 이행(또는 "해결")은 다양한 이행 정보("응답"정보 또는 "해결 정보"라고도 함)가 예를 들어 이행 모듈(124)에 의해 생성/획득되도록 할 수 있다. 후술되는 바와 같이, 이행 정보는 일부의 구현 예에서 자연어 생성기(일부도에서 "NLG")(126)에 제공될 수 있으며, 자연어 생성기(126)는 이행 정보에 기초하여 자연어 출력을 생성할 수 있다.
- [0070] 이행(또는 "해결") 정보는 다양한 방식으로 의도를 이행(또는 "해결")할 수 있기 때문에 다양한 형태를 취할 수 있다. 사용자가 "'샤이닝' 야외 촬영은 어디에서 촬영 되었나요?"와 같은 순수한 정보를 요청한다고 가정한다. 사용자의 의도는 예를 들어 의도 매칭기(135)에 의해 검색 질의로 결정될 수 있다. 검색 질의의 의도 및 내용은, 도 1에 도시된 바와 같이, 응답 정보를 위해 문서 및/또는 다른 데이터 소스(예를 들어 지식 그래프 등)의 코퍼스(corpus)를 검색하도록 구성된 하나 이상의 검색 모듈(150)과 통신할 수 있는 이행 모듈(124)에 제공될 수 있다. 이행 모듈(124)은 검색 질의(예를 들어, 쿼리의 텍스트, 감소된 차원 임베딩 등)를 나타내는 데이터를 검색 모듈(150)에 제공할 수 있다. 검색 모듈(150)은 GPS 좌표와 같은 응답 정보 또는 "Timberline Lodge, Mt. Hood, Oregon"과 같은 보다 명시적인 다른 정보를 제공할 수 있다. 이 응답 정보는 이행 모듈(124)에 의해 생성된 이행 정보의 일부를 형성할 수 있다.
- [0071] 추가로 또는 대안적으로, 이행 모듈(124)은, 예를 들어, 자연어 이해 모듈(135)로부터, 사용자의 의도 및 사용자에 의해 제공되거나, 다른 수단(예를 들어, 사용자의 GPS 좌표, 사용자 선호도, 등)을 사용하여 결정된 임의



의 슬롯 값을 수신하도록 구성될 수 있다. 그리고 응답 동작을 트리거한다. 응답 동작은 예를 들어 상품/서비스 주문, 타이머 시작, 리마인더 설정, 전화 통화 시작, 미디어 재생, 메시지 전송 등을 포함할 수 있다. 이러한 구현에서, 이행 정보는 이행과 연관된 슬롯 값, 확인 응답(일부 경우에 미리 결정된 응답으로부터 선택될 수 있음) 등을 포함할 수 있다.

[0072] 자연어 생성기(126)는 다양한 소스로부터 획득된 데이터에 기초하여 자연어 출력(예를 들어, 인간의 음성을 모방하도록 설계된 단어/구문)을 생성 및/또는 선택하도록 구성될 수 있다. 일부의 구현 예에서, 자연어 생성기(126)는 입력으로서 의도의 이행과 관련된 이행 정보를 수신하고 이행 정보에 기초하여 자연어 출력을 생성하도록 구성될 수 있다. 추가적으로 또는 대안적으로, 자연어 생성기(126)는 서드 파티 애플리케이션(예를 들어, 필요한 슬롯)과 같은 다른 소스로부터 정보를 수신할 수 있으며, 이는 사용자를 위한 자연어 출력을 구성하는 데 사용할 수 있다.

[0073] 도 2는 자동화된 어시스턴트(120)의 클라이언트부(108)가 사용자, 자동화된 어시스턴트(120)(그림 1에서 "자동 어시스턴트 클라이언트"라고도 함) 및 서드 파티 컴퓨팅 서비스(140) 사이의 다중 모드 상호 작용을 용이하게 하도록 구성될 수 있는 방법의 일례를 개략적으로 도시한다. 도 2에서, 클라이언트부(108)는 하나 이상의 라이브러리(242)를 포함할 수 있다. 예를 들어 자동화된 어시스턴트(120)(예를 들어, lib assistant), 캐스트 컴패니언 라이브러리("CCL") 및/또는 클라이언트부(108)의 몰입형 엔진(244)과의 데이터 교환을 용이하게 하는 다른 라이브러리를 포함할 수 있다.

[0074] 몰입형 엔진(244)은 HTML5(HTML, CSS 및 JavaScript 포함), CSS 카드 등과 같은 다양한 다른 기술을 구현하여 서드 파티 컴퓨팅 서비스에 참여하는 사용자에게 시각적 출력을 제공할 수 있다(그렇게 함으로써, 보다 몰입도 높은 사용자 경험을 제공할 수 있다). 다양한 구현에서, 몰입형 엔진(244)은 자바 스크립트 컨트롤러(246)(또는 다른 유사한 컨트롤러) 및 브라우저(248)와 같은 다양한 컴포넌트를 포함할 수 있다. 다양한 구현에서, 임베디드 브라우저(248)는 예를 들어 클라이언트 장치(106)와 연관된 화면에서 GUI로 렌더링될 수 있다. GUI는 사용자, 자동화된 어시스턴트(120) 및 서드 파티 컴퓨팅 서비스(140)를 포함하는 인간 대 컴퓨터 대화와 관련된 정보를 포함할 수 있다. 예를 들어, 일부의 구현 예에서, 서드 파티 컴퓨팅 서비스(140)는, 일 예로 자동화된 어시스턴트(120)의 서버부(119)를 통해, 시각적 표현을 위해 몰입형 엔진(244)에 의해 제공되는 HTML, XML, CSS 등과 같은 마크 업 언어 데이터를 임베디드 브라우저(248)에 제공할 수 있다. 임베디드 브라우저(248)는 웹 뷰(예를 들어, Android 운영 체제 용, iOS 운영 체제 용), iframe 등과 같은 다양한 다른 기술을 사용하여 구현될 수 있다.

[0075] 도 2의 화살표를 따라, 먼저 사용자는 클라이언트부(108)에 구두 자유 형식 입력을 제공할 수 있다. 이 입력은 타이프되거나 음성으로 이루어질 수 있다. 후자의 경우 음성 입력을 분석하여 텍스트 음성 인식 출력을 생성할 수 있다. 어느 경우이든, 텍스트 데이터와 같은 자유 형식 입력을 나타내는 데이터는 클라이언트부(108)에 의해 서버부(119)에 제공될 수 있다. 일부의 구현 예에서, 서버부(119)는 이 텍스트 데이터에 기초하여 사용자의 의도를 결정하려고 시도할 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, 서버부(119)는 서드 파티 컴퓨팅 서비스(140)에 텍스트 데이터를 중계할 수 있다. 따라서 서드 파티 컴퓨팅 서비스(140)가 사용자의 의도를 확인하려고 시도할 수 있다.

[0076] 서버부(119) 또는 서드 파티 컴퓨팅 서비스(140)에 의해 사용자의 의도가 확인되면, 서드 파티 컴퓨팅 서비스(140)는, 일 예로 응답 동작을 취함으로써(예를 들어, 가열 시작) 사용자의 의도를 해결할 수 있다. 추가적으로, 서드 파티 컴퓨팅 서비스(140)와 관련된 하나 이상의 대화 상태 머신은 다양한 상태 사이에서 전환될 수 있다. 예를 들어, 시각적 대화 상태 머신은 하나의 시각적 대화 상태에서 다른 시각적 대화 상태로 전환될 수 있고, 구두 대화 상태 머신은 또한 하나의 언어 대화 상태에서 다른 상태로 전환될 수 있다.

[0077] 상술한 바와 같이, 사용자의 의도가 해결되면 이행 또는 해결 정보가 생성될 수 있다. 이 해결 정보는, 예를 들어 혼합 페이로드(mixed payload)의 일부로서 서드 파티 컴퓨팅 서비스(140)에 의해 서버부(119)에 제공될 수 있다. 해결 정보는 예를 들어 서버부(119)에 의해 사용되어 자연어 출력 생성과 같은 다양한 동작을 수행할 수 있다. 일부의 구현 예에서, 해결 정보 외에, 혼합 페이로드는 마크 업 언어(예 : HTML, XML) 또는 특정 방식(예 : 확대 / 축소, 아래로 스크롤, 특정 아이템에 포커스 등)으로, 기존 GUI와 상호 작용하는 명령과 같은 시각적 대화 상태로 전환되었음을 나타내는 데이터를 포함할 수 있다. 특정 방식으로 기존 GUI와 상호 작용하기 위한 명령의 제공을 통해, 서드 파티 컴퓨팅 서비스(140)는 클라이언트부가 기존 GUI와 상호 작용하도록 명령을 구현하게 함으로써, 상기 클라이언트부(108)를 제어할 수 있다.

[0078] 다양한 구현에서, 서버부(119)는 혼합 페이로드의 적어도 일부를 클라이언트부(108)에 전달할 수 있다. 예를 들

어, 도 2에서, 서버부(119)는 시각적으로 전환된 상태를 나타내는 데이터를 클라이언트부(108)에 제공한다. 예를 들어, 몰입형 엔진(244)을 통해 클라이언트부(108)는 수신된 시각적 상태 데이터에 기초하여 임베디드 브라우저(248)에 의해 렌더링된 GUI를 업데이트할 수 있다. 예를 들어, 혼합 페이로드가 특정 방식으로 기존 GUI와 상호 작용하는 명령을 포함하는 경우, 몰입형 엔진(244)은 명령이 구현되도록 할 수 있다. 그 결과 임베디드 브라우저(248)에 의해 렌더링되는 GUI가 업데이트될 수 있다. GUI의 업데이트된 뷰를 나타내는 데이터를 제공하는 대신, 명령을 사용하여 기존 GUI와 상호 작용할 수 있다. 이 경우 GUI의 업데이트를 위해 클라이언트 장치와 서드 파티 컴퓨팅 서비스(140) 사이에서 전송되어야 하는 데이터가 감소될 수 있다.

[0079] 도 3 및 4는 본 명세서에 설명된 기술이 특정 시나리오에서 어떻게 사용될 수 있는지의 일례를 도시한다. 사용자(301)는 자동화된 어시스턴트(120)와 연결하기 위해 독립 대화형 스피커의 형태를 취하는 제1 클라이언트 장치(306A)를 작동시킨다. 특히, 제1 클라이언트 장치(306A)는 자동화된 어시스턴트(120)의 클라이언트부(108)의 인스턴스가 설치될 수 있다. 이 예에서, 제1 클라이언트 장치(306A)는 온 보드 디스플레이 기능이 없다. 그러나 제1 클라이언트 장치(306A)는 스마트 텔레비전의 형태로 제2 클라이언트 장치(306B)와 네트워크 통신한다. 일부의 구현 예에서, 스마트 텔레비전은, 예를 들어 텔레비전의 USB 또는 HDMI 슬롯에 삽입되는 무선 통신 기능(예를 들어, Wi-Fi, 블루투스)을 갖는 "스마트 동글"에 의해 "스마트"하게 될 수 있다. 추가적으로 또는 대안적으로, 텔레비전은, 자동화된 어시스턴트(120)의 클라이언트부(108)의 다른 인스턴스를 작동할 수 있게 하는 온 보드 로직을 가지고 있기 때문에 본질적으로 "스마트"할 수 있다. 일부의 구현 예에서, 클라이언트 장치(306A 및 306B)는 사용자(301)의 온라인 프로필과 연관된 클라이언트 장치의 동일한 조정된 "에코 시스템(eco system)"의 일부일 수 있다. 일부의 구현 예에서, 제1 클라이언트 장치(306A)에서 동작하는 자동화된 어시스턴트(120)의 클라이언트부(108)의 인스턴스는 그 GUI를 제2 클라이언트 장치(306B)로 "캐스트(cast)"할 수 있다. 예를 들어, 제1 클라이언트 장치(306A)에서 동작하는 클라이언트부(108)의 인스턴스는, 임베디드 브라우저(248)를 제2 클라이언트 장치(306B)로 캐스팅(casting)할 수 있다.

[0080] 도 3에서, 사용자(301)는 자동 어시스턴트(120)와 함께 자동 어시스턴트(120)와 결합하여 자동 온도 조절 장치 또는 전등 스위치용 조광기와 같은 전기 장치의 와이어링(wiring)을 용이하게 한다. 전기 장치를 와이어링하기 위한 청각적 및 시각적 안내는 제1 클라이언트 장치(306A)에서 동작하는 자동화된 어시스턴트(120)의 클라이언트부(108)와 상호 작용함으로써 서드 파티 컴퓨팅 서비스에 의해 제공된다. 자동화된 어시스턴트(120)는 제1 클라이언트 장치(306A)를 통해 음성 출력을 제공하고, 사용자(301)에게 잠재적인 와이어링 구성 목록에서 "당신의 와이어링 구성은 무엇입니까?"라고 집에서의 와이어링 구성을 선택하도록 요청한다. 한편, 자동화된 어시스턴트(120)는 서드 파티 컴퓨팅 서비스에 의해 제공되는 GUI가 제2 클라이언트 장치(306B)에서 렌더링되도록 한다. 이 GUI에는 동일한 배선 구성 질문과 A-C 후보 답변 선택인 일련의 잠재적 와이어링 구성이 포함되어 있다. 이러한 후보 답변 선택 A-C는, 제1 클라이언트 장치(306A)를 통해 자동화된 어시스턴트(120)에 의해 소리내어 읽혀질 수도 있고 읽히지 않을 수도 있다. 사용자(301)는 "Not two-wire"라는 발화를 제공함으로써 후보 답변 선택 A를 제거한다. 후보 응답 선택 A가 맞지 않다는 결정에 응답하여, 서드 파티 컴퓨팅 서비스는 와이어링 구성과 연관된 음성 대화 상태 머신 및 시각적 대화 상태 머신 모두를 업데이트한다.

[0081] 대안적으로, 제2 클라이언트 장치(306B)가 터치 스크린을 포함하고 사용자(301)가 후보 답변 선택 A-C로부터 올바른 와이어링 구성을 알고 있다면, 사용자(301)는 그의 선택된 답변을 탭 했을 수 있다. 일부의 구현 예에서, 선택된 답변은 적어도 일시적으로 시각적으로 강조될 수 있다. 여하튼 사용자(301)에 의한 발화는 자동화된 어시스턴트(120)의 클라이언트부(108) 및 서버부(119)와 관련하여 위에서 설명된 다양한 구성 요소를 사용하여 기록되고 처리될 수 있으며, 사용자(301)에 의해 제공되는 답변을 나타내는 데이터는 결국 서드 파티 컴퓨팅 서비스에 제공될 수 있다.

[0082] 한편, 서드 파티 컴퓨팅 서비스는 시각적 대화 상태 머신의 업데이트된 상태에 기초하여, 자동화된 어시스턴트의 서버부(119)에 시각적 대화 상태 데이터(예를 들어, 도 2를 참조하여 설명된 혼합 페이로드의 일부로서)를 제공할 수 있다. 상기 시각적 대화 상태 데이터는 자동 어시스턴트(120)의 클라이언트부(108)에 차례로 제공될 수 있다. 자동화된 어시스턴트(120)의 클라이언트부(108)는 시각적 대화 상태 데이터를 수신하는 것에 응답하여, 제2 클라이언트 장치(306B)에서 렌더링된 GUI를 업데이트하여 제거된 답변 A가 부정확하다는 시각적 표시를 제공한다(예를 들어 취소선 또는 다른 유사한 시각적 주석 사용). 그러나 사용자(301)는 후보 답변 선택 A만을 제거하고 후보 답변 선택 B-C를 선택하지 않았다.

[0083] 이제 도 4를 참조하면, 서드 파티 컴퓨팅 서비스는 구두 대화 상태 머신의 업데이트된 상태에 기초하여, 자동화된 어시스턴트(120)가 "와이어 중 하나가 녹색입니까?"라는 음성 출력을 렌더링하게 하는 해결 데이터를 생성할 수 있다. 사용자(301)는 "예, 와이어 중 하나가 녹색이다."라는 발화를 제공하여 응답할 수 있다. 서드 파티 컴

퓨팅 서비스는, 언어 대화 상태 머신의 업데이트된 상태에 기초하여 사용자(301)가 접지선이 있는 와이어링 구성을 가지고 있다고 결정할 수 있으며, 자동화된 어시스턴트(120)가 음성 출력 "당신의 와이어링 구성은 아마도 접지가 있는 3선일 것이다."를 렌더링하도록 하는 해결 데이터를 생성한다. 후보 응답 선택 B가 아마도 정확하다는 결정에 응답하여, 서드 파티 컴퓨팅 서비스는 그것이 제공하는 잠재적인 와이어링 구성과 연관된 구두 대화 상태 머신 및 시각적 대화 상태 머신 모두를 업데이트할 수 있다. 한편, 서드 파티 컴퓨팅 서비스는 시각적 대화 상태 머신의 업데이트된 상태에 기초하여 자동화된 어시스턴트의 서버부(119)에 시각적 대화 상태 데이터(예를 들어, 도 2를 참조하여 설명된 혼합 페이로드의 일부로서)를 제공할 수 있다. 시각적 대화 상태 데이터는 자동 어시스턴트(120)의 클라이언트부(108)에 차례로 제공될 수 있다. 자동화된 어시스턴트(120)의 클라이언트부(108)는 시각적 대화 상태 데이터를 수신하는 것에 응답하여, 시각적 표시를 제공하기 위해 제2 클라이언트 장치(306B)에서 렌더링된 GUI를 업데이트할 수 있다. 그리고 올바른 와이어링 구성은 아마도 접지가 있는 3선 구성이거나 후보 답변 선택 B(예 : 굵게, 밑줄, 기울임 꼴 또는 기타 글꼴이나 애니메이션을 사용하여 정답을 강조 표시)일 수 있다.

[0084] 또한, 서드 파티 컴퓨팅 서비스는 시각적 대화 상태 머신의 업데이트된 상태에 기초하여 전기 장치를 와이어링하는 데 사용되는 접지 연결이 있는 3선을 포함하는 와이어링 다이어그램의 시각적 표시를 제공할 수 있다. 여기에는 와이어링 구성으로, 3선 와이어링을 위한 서드 파티 컴퓨팅 서비스의 단계별 다이어그램, 접지 구성으로 3선 와이어링을 위한 서드 파티 컴퓨팅 서비스의 비디오 자습서 등이 포함될 수 있다.

[0085] 도 5 및 도 6은 설명된 기술이 특정 시나리오에서 어떻게 사용될 수 있는지에 대한 또 다른 예를 묘사한다. 도 5에서, 태블릿 컴퓨팅 장치의 형태를 취하는 클라이언트 장치(506)는 현재 사용자가 액세스할 수 있고, 모니터링할 수 있으며, 사용자와 상호 작용하는, 그리고 다양한 컴포넌트를 저장 및/또는 온도 조절기에 관련된 다양한 구성들을 저장하는 인터페이스를 사용자에게 제공하는 서드 파티 컴퓨팅 서비스(예 : 도 3 및 도 4의 온도 조절 장치의 배선 구성과 관련된 서드 파티 컴퓨팅 서비스)와 관련된 GUI를 터치 스크린(570)에 표시하고 있다. 이 예에서 GUI는 서드 파티 컴퓨팅 서비스에서 제공하는 다양한 구성 요소 및/또는 구성과 관련된 일련의 타일(556<sub>1-6</sub>)(또는 카드)을 포함하고 사용자가 클라이언트 장치(506)에서 서드 파티 컴퓨팅 서비스에 의해 제공되는 실행 가능한 콘텐츠와 상호 작용할 수 있도록 한다. 첫 번째 타일(556<sub>1</sub>)은 사용자의 집에서 온도 조절기에 의해 감지된 현재 온도와 관련된다. 두 번째 타일(556<sub>2</sub>)는 사용자의 현재 계정 설정과 관련된다. 세 번째 타일(556<sub>3</sub>)은 사용자의 과거, 현재 및 미래의 온도 조절기 사용과 관련된다. 네 번째 타일(556<sub>4</sub>)는 사용자가 온도 조절기의 작동과 관련된 지원을 위해 서드 파티 컴퓨팅 서비스를 호스팅하는 서드 파티에게 연락할 수 있게 한다. 다섯 번째 타일(556<sub>5</sub>)은 서드 파티 컴퓨팅 서비스에 의해 제공되고 사용자와 상호 작용할 수 있는 온도 조절기의 구성(도 3 및 도 4에서 적절하게 와이어링된 후)과 관련된다.

[0086] 도 5에서 사용자는 "Let's have a look at the bottom tile."라는 발화를 제공한다. 이 발언은 사용자의 의도를 나타내는 데이터가 서드 파티 컴퓨팅 서비스로 전달될 때까지, 앞서 설명된 자동화된 어시스턴트(120)의 다양한 구성 요소를 사용하여 기록되고 처리될 수 있다. 서드 파티 컴퓨팅 서비스는 따라서 시각적 대화 상태 머신 및 구두 대화 상태 머신을 업데이트할 수 있다. 특히, 도 6에 도시된 바와 같이, 시각적 대화 상태 머신은 다섯 번째 타일(556<sub>5</sub>)를 줌 인(즉, 확대)하고, 효과적으로 장치 구성을 시각적 및/또는 구두 대화 상태 머신들의 초점으로 만드는 새로운 상태로 전환된다. 추가적으로, 서드 파티 컴퓨팅 서비스는 터치 스크린(570) 상에 디스플레이하기 위한 복수의 장치 구성 옵션을 제공할 수 있다. 결과적으로 도 6에서 사용자가 "Let's do the bottom one"이라는 추가 발화를 제공하면, 서드 파티 컴퓨팅 서비스는 "bottom one"이라는 용어를 장치 구성의 "온도 설정"으로 해석할 수 있다. 그리고 주변 실내 온도, 실외 온도 또는 다른 사용자가 자주 설정하는 온도를 기반으로 현재 온도에 대한 추가 정보를 제공하거나 사용자에게 온도를 제안할 수 있다. 클라이언트 장치(506)를 통한 서드 파티 컴퓨팅 서비스는 사용자에게 특정 설정 온도를 청각적 또는 촉각적으로 입력하도록 요청할 수 있다. 또한, 사용자는 "저장", "구성 저장"등의 가청 입력을 제공하여 특정 장치 구성을 저장할 수 있다. 또는 사용자는 터치 스크린(570)에서 "장치 구성 저장"을 탭하여 촉각 입력을 제공할 수 있다.

[0087] 여기에 설명된 기술은 서드 파티 개발자가 제공하는 컴퓨팅 서비스 또는 에이전트로 제한되지 않는다. 일부의 구현 예에서, 본 명세서에 설명된 기술은 자동화된 어시스턴트(120) 자체에 의해, 예를 들어, 자동화된 어시스턴트(120)와 관련되거나 심지어 그 일부를 형성하는 "퍼스트 파티(first party) 컴퓨팅 서비스"로서 사용될 수 있다. 도 7은 본 명세서에 설명된 기술이 퍼스트 파티 컴퓨팅 서비스에 의해 사용되는 방법의 예를 도시한다. 도 7에서, 클라이언트 장치(706)는 터치 스크린 디스플레이부(770)의 형태로 디스플레이 기능을 갖는 독립형 대화형 스피커의 형태를 가질 수 있다. 이 예에서, 디스플레이부(770)는 사용자(도시되지 않음)에 대한 잠재적 인



관련성의 다양한 정보를 포함하는 GUI를 렌더링한다. 가장 관련성이 높은 것은 사용자가 "축구에서 조셉을 45분 안에 데리러 가야한다."는 리마인더(reminder)이다. 다양한 구현에서, 이 리마인더는 예를 들어 디스플레이부(770)상에서 및/또는, 클라이언트 장치(706)를 포함하는 클라이언트 장치들이 공동작용하는 에코시스템(ecosystem)의 다른 클라이언트 장치들에서 렌더링되는 카드로서, 사용자에게 자동으로 푸시될 수 있다.

[0088] 이 알림 아래에는 "모든 리마인더 보기(SEE ALL REMINDERS)"라는 "실행 가능한 텍스트"가 표시된다. 다양한 구현에서 사용자는 실행 가능한 텍스트를 선택할 수 있다. 예를 들어 탭하거나 사용자가 실행 가능한 텍스트와 관련된 문구 - 이 예에서는 "모든 리마인더 보기"-를 말할 수 있다. 어느 경우이나 자동화된 어시스턴트(120)에 의해 디스플레이부(770)에 리마인더들을 보여주는 새로운 GUI가 렌더링될 수 있다. 사용자의 관점에서 이 실행 가능한 텍스트는 하이퍼 링크와 유사하게 작동하는 것처럼 보일 수 있다. 그러나 일부의 구현 예에서는 사용자가 이 실행 가능한 텍스트를 탭할 때 HTTP 명령이 HTTP 서버로 전송되어 디스플레이부(770)에 새 웹 페이지를 로드하도록 트리거하지 않는다. 오히려, 사용자의 탭을 나타내는 데이터가 자동화된 어시스턴트(120)의 서버부(119)에 제공될 수 있으며, 이는 사용자의 탭을 나타내는 데이터를 컴퓨팅 서비스(퍼스트 파티 또는 서드 파티)에 제공할 수 있다. 컴퓨팅 서비스는, 예를 들어, 구두 및/또는 시각적 대화 상태 머신을 업데이트하고, 시각적 대화 상태 데이터를 서버부(119)를 통해 자동화된 어시스턴트(120)의 클라이언트부(108)에 제공함으로써, 데이터를 해석(예를 들어, 모든 리마인더를 보기 위한 의도로서)하고 응답할 수 있다.

[0089] 일부의 구현 예에서, 자동화된 어시스턴트(120)는 먼저 호출을 요구하지 않아도 실행 가능한 텍스트에 대응하는 임의의 발화를 "듣도록"구성될 수 있다. 달리 말하면, 실행 가능한 텍스트가 디스플레이부에 제시될 때, 자동화된 어시스턴트(120)는 전통적인 호출 문구(예를 들어, "OK Assistant") 및 표시된 실행 가능한 텍스트와 관련된 발화 모두를 들을 수 있다. 따라서 도 7의 GUI와 마주했을 때, 사용자는 전통적인 호출 문구를 사용하여 자동화된 어시스턴트(120)를 호출할 수 있다. 또한 "모든 리마인더 보기"라는 실행 가능한 텍스트를 말(또는 토크)함으로써 자동화된 어시스턴트(120)를 호출할 수 있다.

[0090] 일부의 구현 예에서, 자동화된 어시스턴트(120)와의 음성 참여에 의해 트리거될 수 있는 다른 작업을 트리거하도록 작동 가능한 추가 타일 또는 카드가 제공될 수 있다. 예를 들어, 제1 타일(756<sub>1</sub>)은 자동화된 어시스턴트(120)가 구두 대화 상태 머신을 업데이트하고 날씨를 설명하는 가청 출력을 제공하도록 동작 가능하다. 사용자가 "다음 주는 어때?"와 같은 모호한 질문을 했다고 가정한다. ("날씨"의 현재 의도가 활성화될 수 있는) 구두 대화 상태 머신의 현재 구두 대화 상태에 기초하여, 자동화된 어시스턴트(120)는 모호한 질의를 "다음 주 날씨가 어때?"와 같은 것으로 해결하고 그에 따라 응답할 수 있다. 또는 "날씨 알려줘"라는 문구는 사용자가 동일한 목표를 달성하기 위해 말(또는 토크)할 수 있는 실행 가능한 텍스트 일 수 있다.

[0091] 도 7의 다른 카드(756<sub>2</sub>)는 예를 들어, 클라이언트 장치(706)의 스피커(도시되지 않음)를 통해 청각적으로 및/또는 디스플레이부(770)를 통해 시각적으로 사용자의 일정을 제시하도록, 사용자에게 의해 동작 가능하다. 다시 한번, 예를 들어 따옴표로 묶여 있는 "What's on my schedule?" 이라는 텍스트는, 사용자가 동일한 목표를 달성하기 위해 카드 756<sub>2</sub>를 토크하는 대신 말할 수 있는 실행 가능한 텍스트 일 수 있다. 일부의 구현 예에서, 사용자가 카드 756<sub>2</sub>를 토크하면, 사용자의 일정이 자동화된 어시스턴트(120)에 의해 디스플레이부(770)에 표시될 수 있다. 반면, 사용자가 실행 가능한 텍스트 "내 일정에 뭐가 있지?" 라고 말하면 자동화된 어시스턴트(120)는 사용자의 일정을 음성으로 출력할 수 있다.

[0092] 도 8은 본 명세서에 개시된 구현에 따른 예시적인 방법(800)을 예시하는 흐름도이다. 편의상, 흐름도의 동작은 동작을 수행하는 시스템을 참조하여 설명한다. 이 시스템은 자동화된 어시스턴트(120)를 구현하는 컴퓨팅 시스템의 하나 이상의 컴포넌트와 같은 다양한 컴퓨터 시스템의 다양한 컴포넌트를 포함할 수 있다. 더욱이, 방법(800)의 동작이 특정 순서로 도시되지만, 이것은 제한하는 것을 의미하지 않는다. 하나 이상의 작업을 다시 정렬하거나 생략하거나 추가할 수 있다. 일부의 구현 예에서, 방법(800)의 동작은 서드 파티 컴퓨팅 서비스(140)를 동작시키는 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있지만, 이것이 필수는 아니다.

[0093] 블록 802에서, 시스템은, 예를 들어, 자동화된 어시스턴트(예를 들어, 서버부(119))로부터 하나 이상의 프로세서에 의해 적어도 부분적으로 구현된(퍼스트 파티 또는 서드 파티) 컴퓨팅 서비스에서, 사용자와 자동화된 어시스턴트 간의 인간 대 컴퓨터 대화 세션의 일부로서 자동화된 어시스턴트와 통신하는 컴퓨팅 장치(예를 들어, 클라이언트 장치(106, 306, 506, 706))에서 사용자 의도를 나타내는 데이터를 수신할 수 있다. 예를 들어, 데이터는 사용자가 제공 한 발화에 기초하여 생성된 음성 인식 출력을 포함할 수 있다. 추가적으로 또는 대안적으로, 일부의 구현 예에서, 데이터는 자동화된 어시스턴트(120)의 서버부(119)(특히, 의도 매칭기(135))에 의해 결정

되는 의도를 포함할 수 있다. 일부의 구현 예에서, 데이터는 컴퓨팅 서비스에 의해 요구되거나 요구되지 않는 하나 이상의 슬롯 값을 포함할 수 있다. 동작의 일 예로서, 자동화된 어시스턴트(120)와 인간 대 컴퓨터 대화 세션에 참여한 사용자는 디스플레이부에 문서가 제시될 수 있다. 사용자는 문서를 아래로 스크롤하여 예를 들어 다음 페이지로 스크롤하라는 명령(음성 및/또는 화면 아래로 스와이프)을 제공할 수 있다.

[0094] 블록 804에서, 시스템은 예를 들어 컴퓨팅 서비스를 통해 해결 정보를 생성하려는 사용자의 의도를 해결할 수 있다. 예를 들어, 실시 예에서, 컴퓨팅 서비스는 표시할 문서의 "다음" 부분(예를 들어 문서의 이전에 표시된 부분 아래 부분)을 결정할 수 있다. 컴퓨팅 서비스는 클라이언트 장치에서 동작하는 자동화된 어시스턴트(120)의 클라이언트부(108)가 GUI를 변경하게 하도록 설계된 명령(예 : 임베디드 브라우저(248)에서 렌더링)을 생성할 수 있으며, 그렇게 함으로써 문서의 다음 부분을 표시할 수 있다.

[0095] 블록 806에서, 예를 들어 컴퓨팅 서비스를 통해, 시스템은 인간 대 컴퓨터 대화 세션과 관련된 컴퓨팅 서비스를 위해 유지되는 디스플레이 컨텍스트 또는 시각적 대화 상태 머신을 업데이트할 수 있다. 다양한 구현 예에서, 블록 806의 업데이트는 블록 802에서 결정된 의도 중 적어도 하나에 적어도 부분적으로 기초할 수 있다. 그리고 블록 804에서 해결 정보가 생성될 수 있다.

[0096] 블록 808에서, 시스템은, 예를 들어 컴퓨팅 서비스를 통해, 업데이트된 디스플레이 컨텍스트 또는 시각적 대화 상태 머신을 나타내는 데이터를 자동화된 어시스턴트(120), 특히 서버부(119)에 제공할 수 있다. 다양한 구현 예에서, 디스플레이 컨텍스트를 나타내는 데이터는 서버부(119)에 의해 컴퓨팅 장치(예를 들어, 106, 306, 506, 706)에 제공될 수 있다. 다양한 구현 예에서, 디스플레이 컨텍스트를 나타내는 데이터는 컴퓨팅 장치에서 실행되는 어시스턴트 애플리케이션(예 : 클라이언트부 108)이, 사용자와 어시스턴트 애플리케이션의 그래픽 사용자 인터페이스 간의 비접촉 상호 작용을 트리거하도록 할 수 있다. 동작 예를 다시 참조하면, 사용자가 아무것도 터치할 필요 없이 문서의 다음 부분으로 "스크롤"되도록 디스플레이부가 변경될 수 있다.

[0097] 도 9는 본 명세서에 개시된 구현에 따른 예시적인 방법(900)을 예시하는 흐름도이다. 편의상, 흐름도의 동작은 동작을 수행하는 시스템을 참조하여 설명한다. 이 시스템은 자동화된 어시스턴트(120)를 구현하는 컴퓨팅 시스템의 하나 이상의 컴포넌트와 같은, 다양한 컴퓨터 시스템의 다양한 컴포넌트를 포함할 수 있다. 더욱이, 방법(900)의 동작이 특정 순서로 도시되지만, 이것은 제한하는 것을 의미하지 않는다. 하나 이상의 작업을 다시 정렬하거나 생략하거나 추가할 수 있다.

[0098] 블록 902에서, 시스템은 디스플레이부(예 : 306B, 570, 770)를 갖는 클라이언트 장치(예 : 106, 306, 506, 706)에서 자동화된 어시스턴트(120)의 클라이언트부(108)를 구현할 수 있다. 블록 904에서, 시스템은 하나 이상의 서버 컴퓨팅 장치에서 자동화된 어시스턴트(120)의 서버부(119)를 구현할 수 있다. 다양한 구현에서, 사용자는 사용자, 자동화된 어시스턴트(120) 및 서드 파티 컴퓨팅 서비스(140) 사이의 인간 대 컴퓨터 대화 세션에 참여하기 위해 클라이언트부(108)와 상호 작용할 수 있다. 다양한 구현에서, 서드 파티 컴퓨팅 서비스(140)는 구두 대화 상태 머신 및 시각적 대화 상태 머신에 따라 사용자와 소통할 수 있다.

[0099] 블록 906에서, 시스템은 예를 들어 클라이언트부(108)에서 자동화 어시스턴트(120)의 서버부(119)를 통해 서드 파티 컴퓨팅 서비스(140)로부터 시각적 대화 상태 머신의 제1 시각적 대화 상태를 나타내는 데이터를 수신할 수 있다. 블록 908에서, 클라이언트부(108)는 제1 시각적 대화 상태에 기초하여, 인간 대 컴퓨터 대화 세션과 관련된 그래픽 사용자 인터페이스를 디스플레이부에 렌더링할 수 있다. 다양한 구현에서, 그래픽 사용자 인터페이스는, 구두 상태 머신이 제1 시각적 대화 상태에 대응하는 제1 구두 대화 상태에서 제2 구두 대화 상태로의 전환을 유발하는 적어도 하나의 그래픽 요소(예 : 버튼, 실행 가능한 텍스트 등)를 포함할 수 있다.

[0100] 블록 910에서, 클라이언트부(108)는 사용자에게 의한 적어도 하나의 그래픽 요소의 동작을 검출할 수 있다. 예를 들어, 사용자는 그래픽 요소를 탭 하거나 명령을 말할 수 있다(예를 들어, 그래픽 요소와 상호 작용하기 위해 실행 가능한 텍스트와 관련된 문구를 발언함). 블록 912에서, 클라이언트부(108)는 적어도 하나의 그래픽 요소(예를 들어, 의도)의 동작을 나타내는 데이터를 서버부(119)에 제공할 수 있다. 블록 914에서, 서버부(119)는 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터를 서드 파티 컴퓨팅 서비스(140)에 제공할 수 있다. 다양한 구현 예에서, 적어도 하나의 그래픽 요소의 동작을 나타내는 데이터는 서드 파티 컴퓨팅 서비스(140)가 제1 구두 대화 상태에서 제2 구두 대화 상태로의 전환을 유발할 수 있다.

[0101] 도 10은 본 명세서에 설명된 기술의 하나 이상의 양상을 수행하기 위해 선택적으로 이용될 수 있는 예시적인 컴퓨팅 장치(1010)의 블록도이다. 일부의 구현 예에서, 클라이언트 컴퓨팅 장치, 사용자 제어 리소스 모듈(130) 및/또는 다른 구성 요소(들) 중 하나 이상은 예시적인 컴퓨팅 장치(1010)의 하나 이상의 구성 요소를 포함할 수

있다.

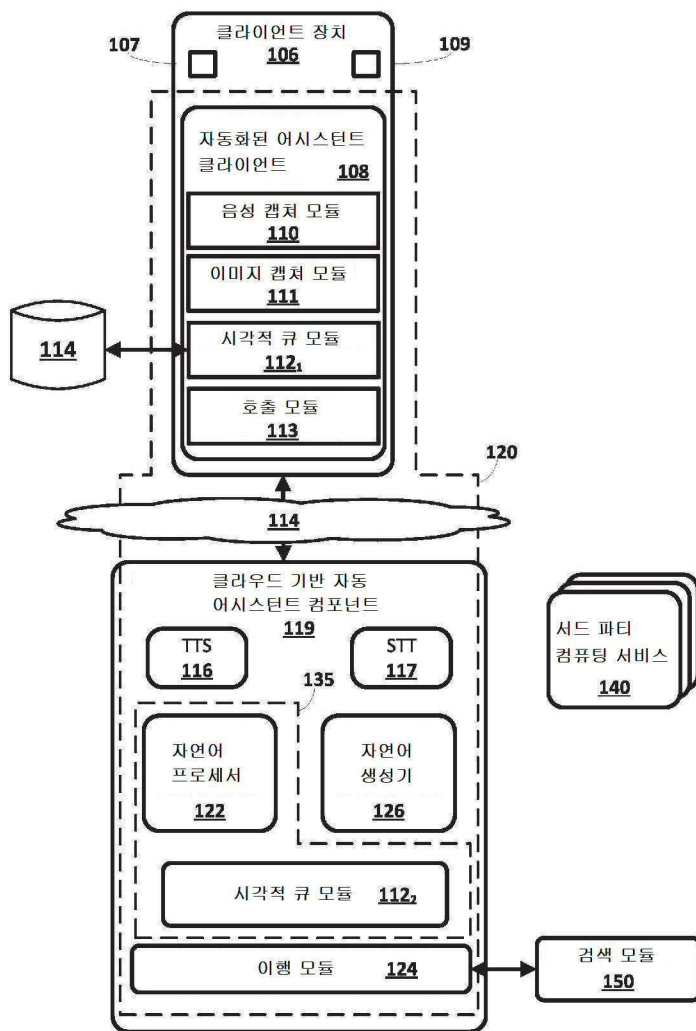
- [0102] 컴퓨팅 장치(1010)는 일반적으로 버스 서브 시스템(1012)을 통해 다수의 주변 장치와 통신하는 적어도 하나의 프로세서(1014)를 포함한다. 이러한 주변 장치는 예를 들어 메모리 서브 시스템(1025) 및 파일 저장 서브 시스템(1026), 사용자 인터페이스 출력 장치(1020), 사용자 인터페이스 입력 장치(1022) 및 네트워크 인터페이스 서브 시스템(1016)을 구비하는 저장 서브 시스템(1024)을 포함할 수 있다. 입력 및 출력 장치는 컴퓨팅 장치(1010)와 사용자 상호 작용을 허용한다. 네트워크 인터페이스 서브 시스템(1016)은 외부 네트워크에 대한 인터페이스를 제공하고 다른 컴퓨팅 장치의 대응하는 인터페이스 장치에 연결된다.
- [0103] 사용자 인터페이스 입력 장치(1022)는 키보드, 마우스, 트랙볼, 터치 패드 또는 그래픽 태블릿과 같은 포인팅 장치, 스캐너, 디스플레이부에 통합된 터치 스크린, 음성 인식 시스템과 같은 오디오 입력 장치, 마이크 및/또는 기타 다른 유형의 입력 장치들을 포함할 수 있다. 일반적으로, "입력 장치"라는 용어의 사용은 모든 가능한 유형의 장치 및 컴퓨팅 장치(1010) 또는 통신 네트워크에 정보를 입력하는 방법을 포함하도록 의도된다.
- [0104] 사용자 인터페이스 출력 장치(1020)는 디스플레이 서브 시스템, 프린터, 팩스 머신, 또는 오디오 출력 장치와 같은 비 시각적 디스플레이를 포함할 수 있다. 디스플레이 서브 시스템은 음극선 관(CRT), 액정 디스플레이(LCD)와 같은 평판 장치, 프로젝션 장치, 또는 가시적 이미지를 생성하기 위한 다른 메커니즘을 포함할 수 있다. 디스플레이 서브 시스템은 또한 오디오 출력 장치를 통해 비시각적 디스플레이를 제공할 수 있다. 일반적으로, "출력 장치"라는 용어의 사용은 모든 가능한 유형의 장치 및 컴퓨팅 장치(1010)로부터 사용자 또는 다른 기기 또는 컴퓨팅 장치로 정보를 출력하는 방법을 포함하도록 의도된다.
- [0105] 저장 서브 시스템(1024)은 상술된 모듈의 일부 또는 전부의 기능을 제공하는 프로그래밍 및 데이터 구조를 저장한다. 예를 들어, 저장 서브 시스템(1024)은 도 8 내지 도 9의 방법들에서 선택된 측면을 수행할 뿐만 아니라, 도 1과 도 2에 묘사된 다양한 구성 요소를 구현하기 위한 로직(logic)을 포함할 수 있다.
- [0106] 이러한 소프트웨어 모듈은 일반적으로 프로세서(1014)에 의해 단독으로 또는 다른 프로세서와 결합하여 실행된다. 저장 서브 시스템(1024)에서 사용되는 메모리(1025)는 프로그램 실행 동안 명령어 및 데이터를 저장하기 위한 메인 랜덤 액세스 메모리(RAM)(1030) 및 고정 명령어가 저장되는 읽기 전용 메모리(ROM)(1032)를 포함하는 다수의 메모리를 포함할 수 있다. 파일 저장 서브 시스템(1026)은 프로그램 및 데이터 파일에 대한 영구 저장을 제공할 수 있으며, 하드 디스크 드라이브, 플로피 디스크 드라이브와 관련된 이동식 미디어, CD-ROM 드라이브, 광학 드라이브 또는 이동식 미디어 카트리지를 포함할 수 있다. 특정 구현 기능을 구현하는 모듈은, 저장 서브 시스템(1024)의 파일 저장 서브 시스템(1026)에 의해, 또는 프로세서(들)(1014)에 의해 액세스 가능한 다른 기기에 저장될 수 있다.
- [0107] 버스 서브 시스템(1012)은 컴퓨팅 장치(1010)의 다양한 구성 요소 및 서브 시스템이 서로 간에 통신할 수 있도록 하는 메커니즘을 제공한다. 버스 서브 시스템(1012)이 단일 버스로서 개략적으로 도시되어 있지만, 버스 서브 시스템의 대안적인 구현은 다중 버스를 사용할 수 있다.
- [0108] 컴퓨팅 장치(1010)는 워크 스테이션, 서버, 컴퓨팅 클러스터, 블레이드 서버, 서버 팜, 또는 임의의 다른 데이터 처리 시스템 또는 컴퓨팅 장치를 포함하는 다양한 유형일 수 있다. 컴퓨터 및 네트워크가 끊임없이 변화하는 특성으로 인해, 도 10에 도시된 컴퓨팅 장치(1010)의 설명은 일부 구현을 설명하기 위한 특정 예로서만 의도된다. 컴퓨팅 장치(1010)의 많은 다른 구성은 도 10에 도시된 컴퓨팅 장치보다 더 많거나 적은 구성 요소를 가질 수 있다.
- [0109] 여기에 설명된 시스템이 사용자에게 대한 개인 정보를 수집 또는 모니터링하거나 개인 및/또는 모니터링되는 정보(예 : 사용자의 소셜 네트워크, 소셜 활동 또는 활동, 직업, 사용자의 선호도 또는 사용자의 현재 지리적 위치에 대한 정보)를 사용할 수 있는 상황에서, 사용자는 프로그램 또는 기능이 사용자 정보를 수집하는지 여부를 제어할 수 있는 기회, 또는 사용자와 더 관련이 있을 수 있는 콘텐츠 서버로부터 콘텐츠를 수신할지 여부 및/또는 방법을 제어할 기회를 제공받을 수 있다. 또한 특정 데이터는 저장 또는 사용되기 전에 하나 이상의 방법으로 처리되어 개인 식별 정보가 제거될 수 있다. 예를 들어, 사용자에게 대한 개인 식별 정보를 확인할 수 없도록 사용자의 신원을 처리하거나, 사용자의 지리적 위치로부터 획득된 지리적 위치 정보(예 : 도시, 우편 번호(zip code) 또는 주(state))를 일반화할 수 있다. 따라서 사용자의 특정 지리적 위치를 확인할 수 없다. 그러므로 사용자는 사용자에게 대한 정보 수집 및/또는 사용 방법을 제어할 수 있다. 예를 들어, 일부의 구현 예에서, 사용자는 예를 들어 비전 센서(107)를 비활성화함으로써 시각적 큐를 검출하려는 어시스턴트 장치에서 옵트 아웃(opt out)할 수 있다.

[0110]

여러 구현이 여기에 설명되고 예시되었지만, 기능의 수행 및/또는, 결과 및/또는 여기에 설명된 하나 이상의 이점을 획득하기 위한 다양한 다른 수단 및/또는 구조가 활용될 수 있다. 그리고 그러한 각각의 변형 및/또는 수정은 여기에 설명된 구현의 범위 내에 있는 것으로 간주된다. 보다 일반적으로, 본 명세서에 설명된 모든 매개 변수, 치수, 재료 및 구성은 예시적인 것이며 실제 매개 변수, 치수, 재료 및/또는 구성은 교시가 사용되는 특정 애플리케이션 또는 애플리케이션들에 따라 달라질 수 있다. 당업자는 단지 일상적인 실험을 사용하여 본 명세서에 설명된 특정 구현에 대한 많은 등가물을 인식하거나 확인할 수 있을 것이다. 따라서 진술한 구현은 단지 예로서 제시된 것이며, 첨부된 청구 범위 및 그 균등 물의 범위 내에서, 구현이 구체적으로 설명되고 청구된 것과 달리 실행될 수 있다는 것을 이해해야 한다. 본 개시 내용의 구현은 본원에 기재된 각각의 개별적인 특징, 시스템, 물품, 재료, 키트 및/또는 방법에 관한 것이다. 또한, 그러한 특징, 시스템, 물품, 재료, 키트 및/또는 방법이 서로 일치하지 않는 경우, 이러한 특징, 시스템, 물품, 재료, 키트 및/또는 방법의 임의의 조합은 본 개시의 범위 내에 포함된다.

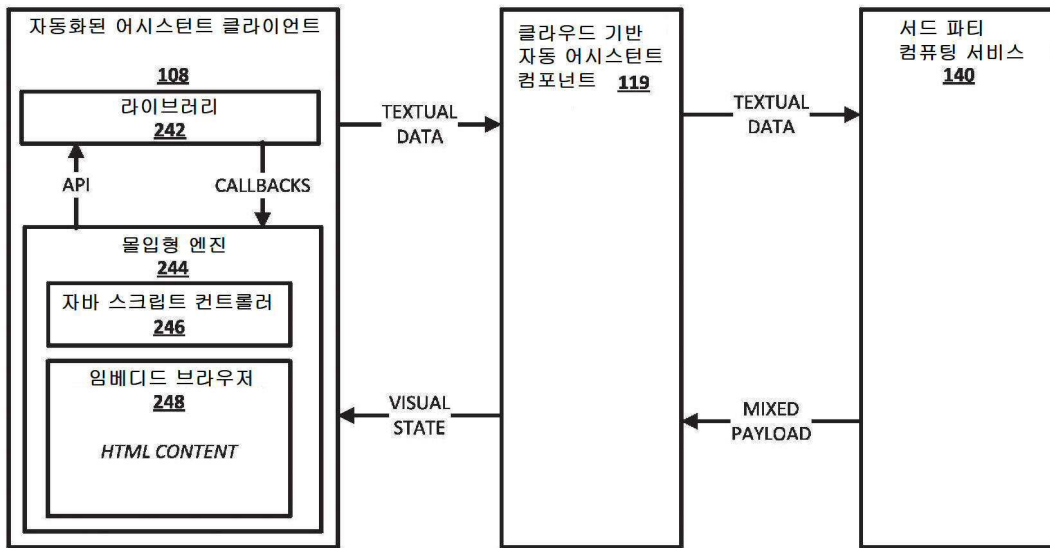
**도면**

**도면1**

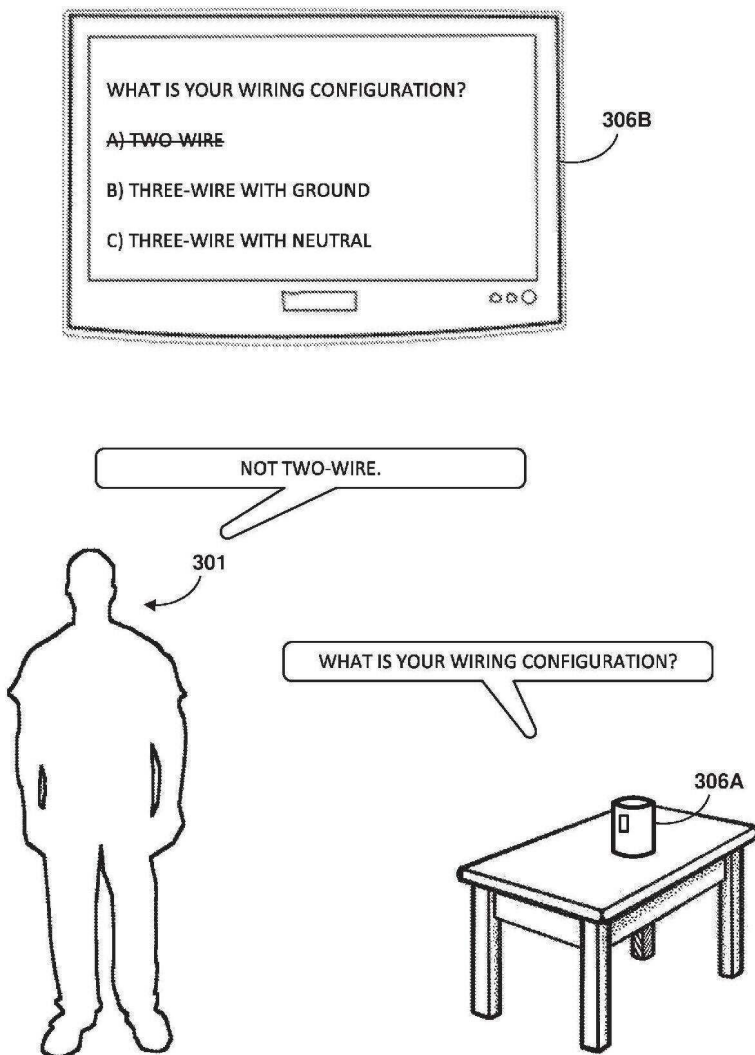




도면2

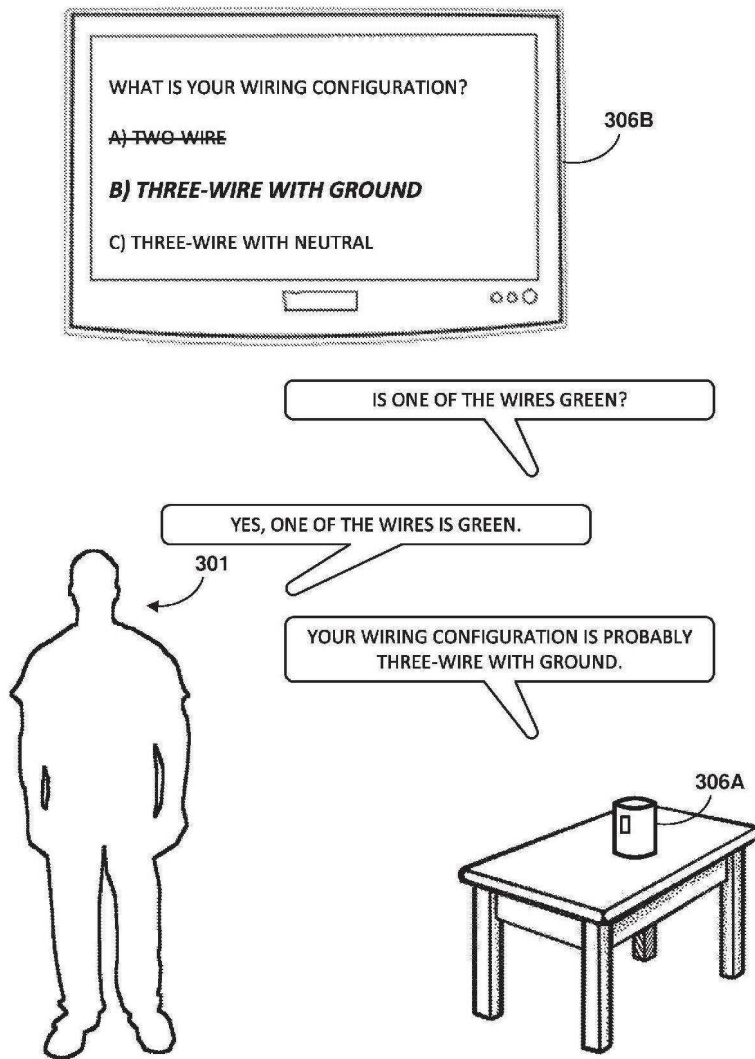


도면3

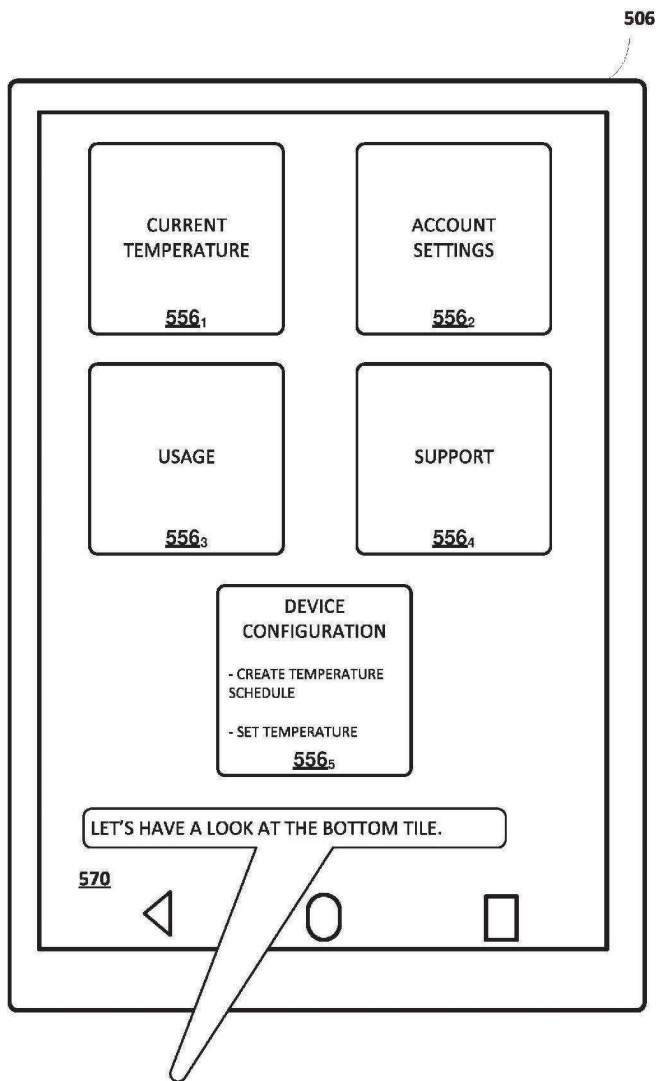




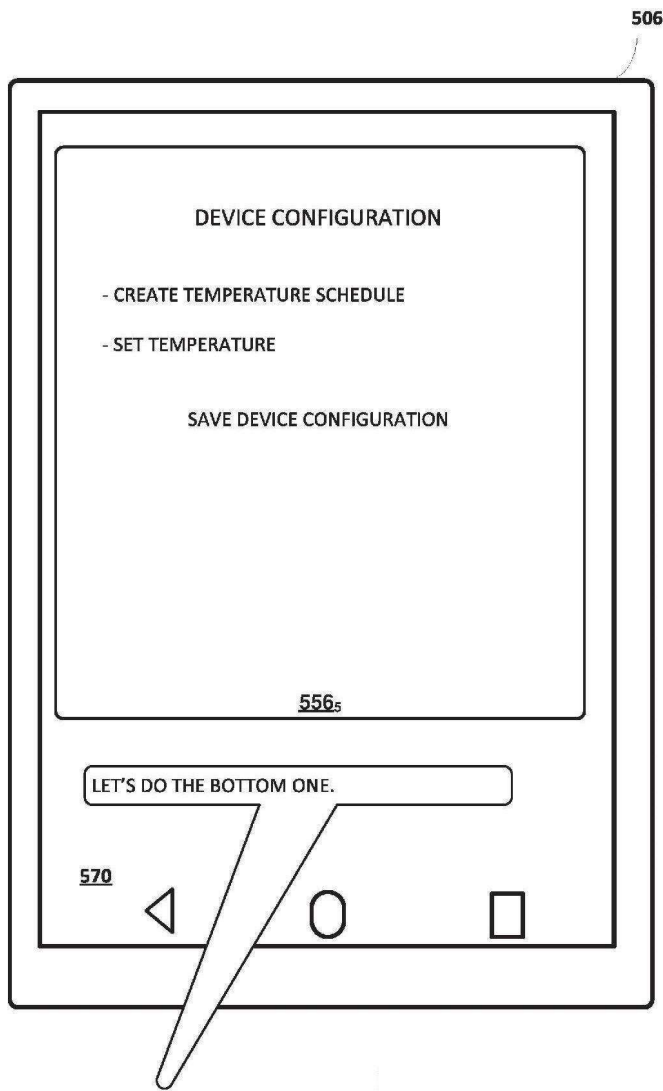
도면4



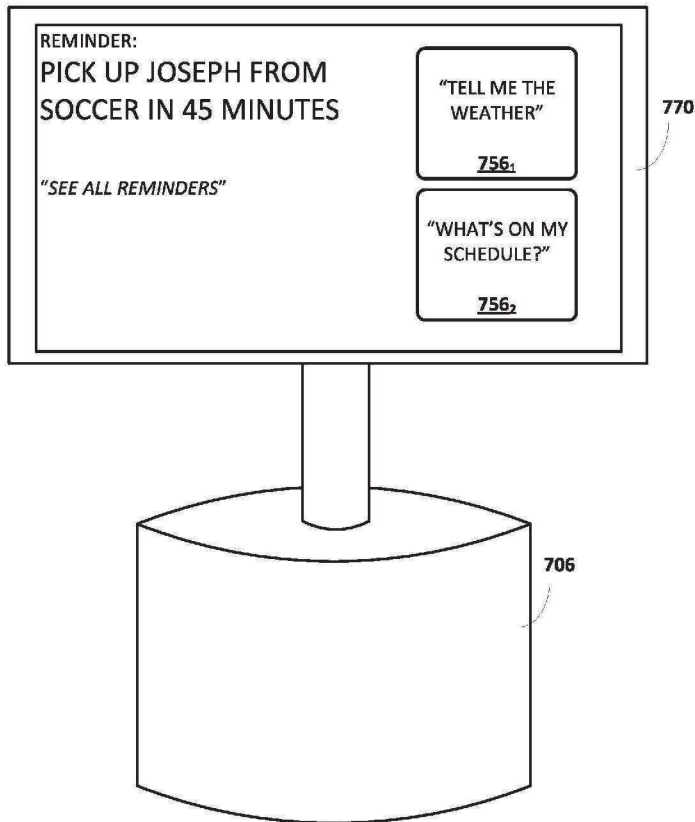
도면5



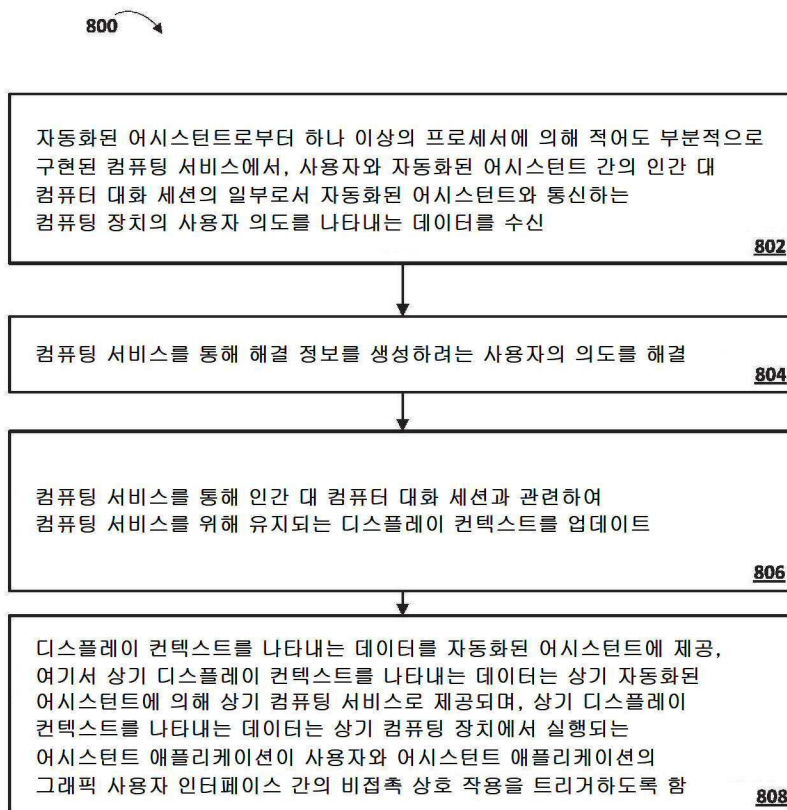
도면6



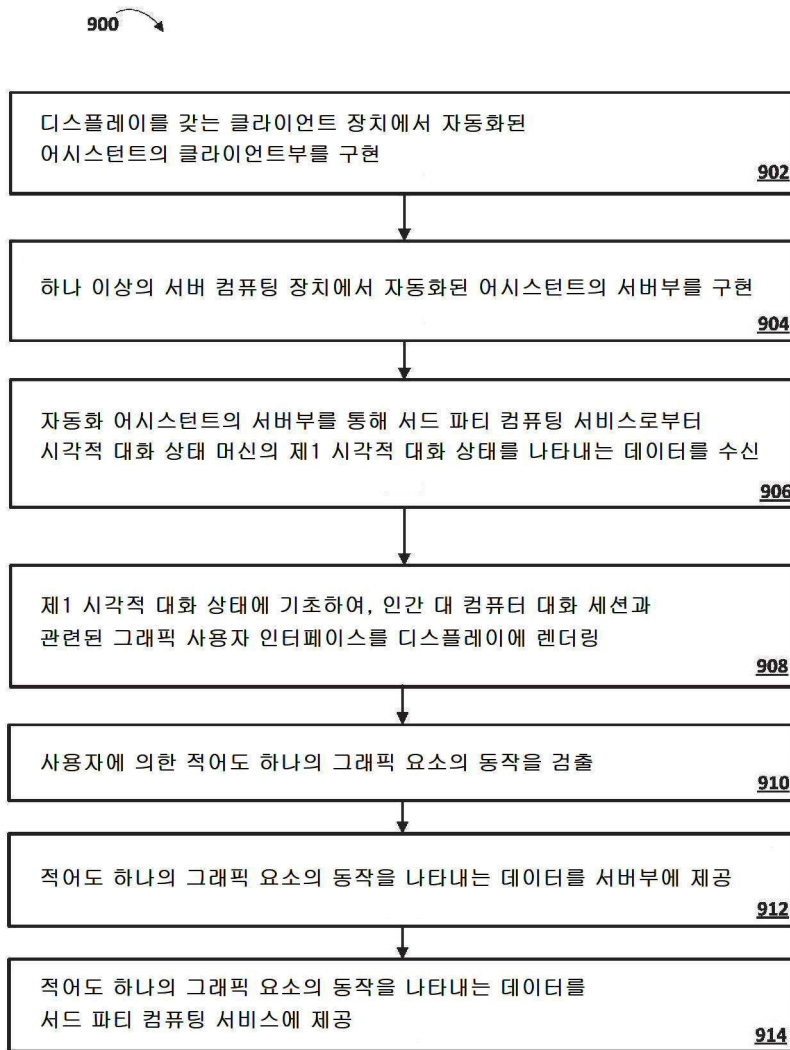
도면7



도면8



도면9



도면10

