



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2024년01월08일  
(11) 등록번호 10-2622356  
(24) 등록일자 2024년01월03일

- (51) 국제특허분류(Int. Cl.)  
G10L 17/06 (2013.01) G06F 16/635 (2019.01)  
G06F 21/32 (2013.01) G06V 40/10 (2022.01)  
G10L 15/07 (2013.01) G10L 15/08 (2006.01)  
G10L 15/22 (2006.01) G10L 15/26 (2006.01)
- (52) CPC특허분류  
G10L 17/06 (2013.01)  
G06F 16/636 (2019.01)
- (21) 출원번호 10-2022-7036565(분할)
- (22) 출원일자(국제) 2018년04월18일  
심사청구일자 2022년10월20일
- (85) 번역문제출일자 2022년10월20일
- (65) 공개번호 10-2022-0148319
- (43) 공개일자 2022년11월04일
- (62) 원출원 특허 10-2021-7035224  
원출원일자(국제) 2018년04월18일  
심사청구일자 2021년10월28일
- (86) 국제출원번호 PCT/US2018/028149
- (87) 국제공개번호 WO 2018/195185  
국제공개일자 2018년10월25일
- (30) 우선권주장  
62/488,000 2017년04월20일 미국(US)  
62/567,372 2017년10월03일 미국(US)
- (56) 선행기술조사문헌  
US20170025124 A1  
(뒷면에 계속)

- (73) 특허권자  
구글 엘엘씨  
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)
- (72) 발명자  
옥템 멜렘  
캐나다 온타리오 엠5에이치 2취4 토론토 리치몬드 스트리트 웨스트 111 12층  
조그리카 타랄 프레딕  
미국 캘리포니아 마운틴 뷰 엠피시어터 파크웨이 1600 (우:94043)  
(뒷면에 계속)
- (74) 대리인  
박장원

전체 청구항 수 : 총 19 항

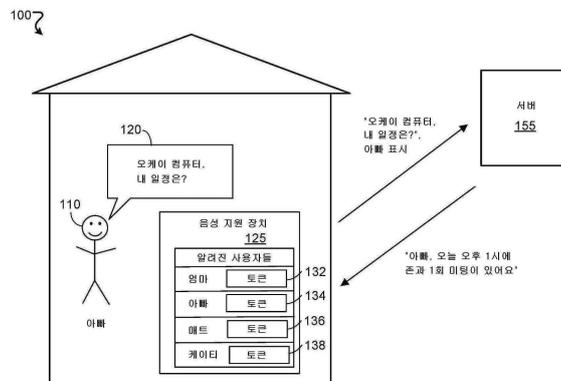
심사관 : 김경완

(54) 발명의 명칭 장치에 대한 다중 사용자 인증

(57) 요약

일부 구현예들에서, 장치의 알려진 사용자들에 대응하는 인증 토큰들이 상기 장에 저장된다. 화자의 발화가 수신된다. 상기 발화는 알려진 사용자들 중 특정 알려진 사용자가 말한 것으로 분류된다. 화자로서 특정 알려진 사용자의 표시 및 발화의 표현을 포함하는 쿼리는 특정 알려진 사용자의 인증 토큰을 사용하여 제공된다.

대표도



(52) CPC특허분류

*G06F 21/32* (2013.01)  
*G06V 40/10* (2022.01)  
*G10L 15/07* (2013.01)  
*G10L 15/08* (2013.01)  
*G10L 15/22* (2013.01)  
*G10L 15/26* (2013.01)  
*G10L 2015/088* (2013.01)

(56) 선행기술조사문헌

US20160174074 A1  
WO2015199813 A1  
KR1020170131585 A  
KR1020170094415 A

(72) 발명자

**헤런디 프누**

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

**차우 푸 셴**

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

**모리노 이그나시오 로페즈**

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

**라자디아샤 살릴**

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

**그루엔스테인 알렉산더 에이치.**

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

**카사도 디에고 멜렌도**

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

## 명세서

### 청구범위

#### 청구항 1

하나 이상의 프로세서에 의해 구현되는 방법으로서,

화자로부터 발화를 수신하는 단계;

상기 화자가 사용자 장치의 알려진 사용자인지 또는 복수의 알려진 사용자에 의해 공유되는 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 단계;

상기 발화가 개인적(personal) 요청에 대응하는지 또는 비-개인적 요청에 대응하는지 여부를 결정하는 단계; 그리고

상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하여 그리고 상기 발화가 비-개인적 요청에 대응한다는 결정에 응답하여:

상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 상기 발화에 대한 응답이 제공되도록 하는 단계, 또는

상기 발화에 응답하여 상기 사용자 장치에 의해 동작(action)이 수행되게 하는 단계를 포함하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

#### 청구항 2

제1항에 있어서, 상기 방법은,

상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하여 그리고 상기 발화가 개인적 요청에 대응한다는 결정에 응답하여:

상기 발화에 대한 응답이 상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 제공될 수 없음을 나타내는 또는 상기 발화에 응답하여 상기 동작이 상기 사용자 장치에 의해 수행될 수 없음을 나타내는 액세스 안됨 메시지(no access message)가 상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 제공되게 하는 단계를 더 포함하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

#### 청구항 3

제2항에 있어서, 상기 방법은,

상기 화자가 상기 사용자 장치의 알려진 사용자라는 결정에 응답하여:

상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 상기 발화에 대한 개인화된(personalized) 응답이 제공되도록 하는 단계 -상기 발화에 대한 개인화된 응답은 상기 화자의 신원(identity)을 기반으로 개인화됨-, 또는

상기 발화에 응답하여 상기 사용자 장치에 의해 개인화된 동작이 수행되도록 하는 단계를 더 포함하며, 상기 수행될 개인화된 동작은 상기 화자의 신원에 기초하여 개인화되는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

#### 청구항 4

제1항에 있어서, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 상기 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 단계는,

상기 사용자 장치에서 로컬로, 그리고 상기 사용자 장치의 복수의 알려진 사용자에 대응하는 인증 토큰에 기초하여, 상기 발화의 화자 식별 특징에 기초하여 상기 발화의 화자를 사용자 장치의 알려진 사용자 또는 상기 사용자 장치의 알려진 사용자가 아닌 것으로 분류하는 단계를 포함하는 것을 특징으로 하는 하나 이상의 프로세서

에 의해 구현되는 방법.

**청구항 5**

제1항에 있어서, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 상기 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 단계는,

상기 발화의 표현 그리고 상기 사용자 장치의 복수의 알려진 사용자에 대응하는 인증 토큰을 적어도 포함하는 쿼리를 생성하는 단계;

상기 쿼리를 원격 서버로 전송하는 단계; 그리고

상기 원격 서버로부터, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 알려진 사용자가 아닌지 여부의 표시를 수신하는 단계를 포함하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

**청구항 6**

제1항에 있어서, 상기 발화가 개인적 요청에 대응하는지 또는 비-개인적 요청에 대응하는지 여부를 결정하는 단계는,

상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보가 상기 응답이 제공되도록 하거나 상기 동작이 수행되도록 하는 데 필요한지 여부를 결정하는 단계; 그리고

상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보가 상기 응답이 제공되도록 하거나 상기 동작이 수행되도록 하는 데 필요하지 않다는 결정에 응답하여 상기 발화가 비-개인적 요청이라고 결정하는 단계를 포함하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

**청구항 7**

제1항에 있어서, 상기 발화가 개인적 요청 또는 비-개인적 요청에 대응하는지 여부를 결정하는 단계는,

상기 발화의 표현을 적어도 포함하는 쿼리를 생성하는 단계;

상기 쿼리를 원격 서버로 전송하는 단계; 그리고

상기 원격 서버로부터, 상기 발화가 개인적 요청 또는 비-개인적 요청에 대응하는지 여부의 표시를 수신하는 단계를 포함하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

**청구항 8**

제1항에 있어서, 상기 발화가 개인적 요청에 대응하는지 또는 비-개인적 요청에 대응하는지를 결정하는 단계는 상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

**청구항 9**

제1항에 있어서, 개인적 요청은 상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보에 의존하고, 비-개인적 요청은 상기 사용자 장치의 복수의 알려진 사용자 중 임의의 사용자의 개인 정보에 의존하지 않는 것을 특징으로 하는 하나 이상의 프로세서에 의해 구현되는 방법.

**청구항 10**

사용자 장치로서,

적어도 하나의 프로세서; 그리고

실행될 때 상기 적어도 하나의 프로세서로 하여금 동작들을 수행하도록 하는 명령어를 저장한 메모리를 포함하며, 상기 동작들은,

화자가 사용자 장치의 알려진 사용자인지 또는 상기 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 동작 -상기 사용자 장치는 복수의 알려진 사용자에 의해 공유됨-;

발화가 개인적(personal) 요청에 대응하는지 또는 비-개인적 요청에 대응하는지 여부를 결정하는 동작; 그리고

상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하여 그리고 상기 발화가 비-개인적 요청에 대응한다는 결정에 응답하여:

상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 상기 발화에 대한 응답이 제공되도록 하는 동작, 또는

상기 발화에 응답하여 상기 사용자 장치에 의해 액션(action)이 수행되게 하는 동작을 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 11**

제10항에 있어서, 상기 동작들은,

상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하여 그리고 상기 발화가 개인적 요청에 대응한다는 결정에 응답하여:

상기 발화에 대한 응답이 상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 제공될 수 없음을 나타내는 또는 상기 발화에 응답하여 상기 액션이 상기 사용자 장치에 의해 수행될 수 없음을 나타내는 액세스 안됨 메시지(no access message)가 상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 제공되게 하는 동작을 더 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 12**

제11항에 있어서, 상기 동작들은,

상기 화자가 상기 사용자 장치의 알려진 사용자라는 결정에 응답하여:

상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 상기 발화에 대한 개인화된(personalized) 응답이 제공되도록 하는 동작 -상기 발화에 대한 개인화된 응답은 상기 화자의 신원(identity)을 기반으로 개인화됨-, 또는

상기 발화에 응답하여 상기 사용자 장치에 의해 개인화된 액션이 수행되도록 하는 동작을 더 포함하며, 상기 수행될 개인화된 액션은 상기 화자의 신원에 기초하여 개인화되는 것을 특징으로 하는 사용자 장치.

**청구항 13**

제10항에 있어서, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 상기 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 동작은,

상기 사용자 장치에서 로컬로, 그리고 상기 사용자 장치의 복수의 알려진 사용자에 대응하는 인증 토큰에 기초하여, 상기 발화의 화자 식별 특징에 기초하여 상기 발화의 화자를 사용자 장치의 알려진 사용자 또는 상기 사용자 장치의 알려진 사용자가 아닌 것으로 분류하는 동작을 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 14**

제10항에 있어서, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 상기 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 동작은,

상기 발화의 표현 그리고 상기 사용자 장치의 복수의 알려진 사용자에 대응하는 인증 토큰을 적어도 포함하는 쿼리를 생성하는 동작;

상기 쿼리를 원격 서버로 전송하는 동작; 그리고

상기 원격 서버로부터, 상기 화자가 상기 사용자 장치의 알려진 사용자인지 또는 알려진 사용자가 아닌지 여부를 표시를 수신하는 동작을 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 15**

제10항에 있어서, 상기 발화가 개인적 요청에 대응하는지 또는 비-개인적 요청에 대응하는지 여부를 결정하는 동작은,

상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보가 상기 응답이 제공되도록 하거나 상기

액션이 수행되도록 하는 데 필요한지 여부를 결정하는 동작; 그리고

상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보가 상기 응답이 제공되도록 하거나 상기 액션이 수행되도록 하는 데 필요하지 않다는 결정에 응답하여 상기 발화가 비-개인적 요청이라고 결정하는 동작을 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 16**

제10항에 있어서, 상기 발화가 개인적 요청 또는 비-개인적 요청에 대응하는지 여부를 결정하는 동작은,

상기 발화의 표현을 적어도 포함하는 쿼리를 생성하는 동작;

상기 쿼리를 원격 서버로 전송하는 동작; 그리고

상기 원격 서버로부터, 상기 발화가 개인적 요청 또는 비-개인적 요청에 대응하는지 여부를 표시를 수신하는 동작을 포함하는 것을 특징으로 하는 사용자 장치.

**청구항 17**

제10항에 있어서, 상기 발화가 개인적 요청에 대응하는지 또는 비-개인적 요청에 대응하는지를 결정하는 동작은 상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하는 것을 특징으로 하는 사용자 장치.

**청구항 18**

제10항에 있어서, 개인적 요청은 상기 사용자 장치의 복수의 알려진 사용자 중 한 명 이상의 개인 정보에 의존하고, 비-개인적 요청은 상기 사용자 장치의 복수의 알려진 사용자 중 임의의 사용자의 개인 정보에 의존하지 않는 것을 특징으로 하는 사용자 장치.

**청구항 19**

실행될 때 적어도 하나의 프로세서로 하여금 동작들을 수행하게 하는 명령어를 저장한 비일시적 컴퓨터 판독가능 저장 매체로서, 상기 동작들은,

화자로부터 발화를 수신하는 동작;

상기 화자가 사용자 장치의 알려진 사용자인지 또는 복수의 알려진 사용자에 의해 공유된 사용자 장치의 알려진 사용자가 아닌지 여부를 결정하는 동작;

상기 발화가 개인적(personal) 요청에 대응하는지 또는 비-개인적 요청에 대응하는지 여부를 결정하는 동작; 그리고

상기 화자가 상기 사용자 장치의 알려진 사용자가 아니라는 결정에 응답하여 그리고 상기 발화가 비-개인적 요청에 대응한다는 결정에 응답하여:

    상기 발화에 응답하여 상기 사용자 장치에서 상기 화자에게 제시하기 위해 상기 발화에 대한 응답이 제공되도록 하는 동작, 또는

    상기 발화에 응답하여 상기 사용자 장치에 의해 액션(action)이 수행되게 하는 동작을 포함하는 것을 특징으로 하는 비일시적 컴퓨터 판독가능 저장 매체.

**발명의 설명**

**기술 분야**

[0001] 본 명세서는 일반적으로 자연어 프로세싱에 관한 것이다.

**배경 기술**

[0002] 음성 지원 장치는 사용자의 발화에 대한 음성 응답을 제공할 수 있다. 이러한 장치는 종종 소위 디지털 어시스턴트 소프트웨어 또는 "디지털 어시스턴트" 또는 "가상 어시스턴트"의 프론트 엔드이다. 이러한 장치는 종종 제한된 양의 음성 인식 기능 자체를 제공하고 추가 리소스를 위해 서버 또는 다른 장치와 통신한다. 또한, 이러한 장치는 종종 사용자를 위해 가청 합성 음성으로 변환되는 (예를 들어, 음성 입력된 쿼리 또는 발화에 응답하여)

서버 또는 다른 장치로부터 데이터를 수신한다. 예를 들어, 사용자가 "오늘 비가 올까요?"라고 말하고 음성 지원 장치가 "하루 종일 맑을 거예요"라고 청각적으로 응답할 수 있다. 사용자는 사용자의 개인 정보와 관련된 발화를 제공할 수 있다. 예를 들어, 사용자는 시스템에 "나의 일정에 있는 내용"을 요청할 수 있으며, 장치는 장치와 관련된 일정상의 약속으로 응답할 수 있다. 그러나, 일반적인 음성 지원 장치는 특히 특정 사용자의 개인 정보에 대한 보안 액세스와 관련하여 다양한 사용자를 적절히 구별하지 못하고, 예를 들어 게스트 사용자에게 의한 음성 지원 장치의 일반적인 사용을 여전히 허용한다.

**발명의 내용**

**해결하려는 과제**

**과제의 해결 수단**

- [0003] 음성 지원 장치는 여러 다른 사용자가 사용할 수 있다. 예를 들어, 주방 조리대에 설치된 음성 지원 장치는 가정에서 "존(John)"과 "제인 도(Jane Doe)"가 사용할 수 있다. 음성 지원 장치의 사용자들은 사용자의 개인 정보와 관련된 질문을 할 수 있다. 예를 들어 John과 Jane Doe는 음성 지원 장치에 각자의 요일 일정을 별도로 요청할 수 있다. John 및 Jane Doe 이외의 사용자(예: "게스트들(guests)")들도 사용자의 개인 정보와 관련이 없는 질문을 음성 지원 장치에 할 수 있다. 예를 들어, 게스트 사용자는 음성 지원 장치에 현재 시간을 물어볼 수 있다.
- [0004] 다수의 상이한 사용자들로부터의 발화를 처리하기 위해, 음성 지원 장치는 발화의 화자를 특정 사용자로서 식별하려고 시도할 수 있고, 성공하면 식별된 특정 사용자에게 대응하는 표시(indication)와 함께 서버에 쿼리를 제공하고, 성공적이지 않으면 식별된 특정 사용자에게 대응하는 표시 없이 서버에 쿼리를 제공한다. 예를 들어, 음성 지원 장치는 "John Doe"의 사전 설정된 음성 패턴과 일치하는 것으로 발화를 인식하는 것에 기초하여 "John Doe"에 의해 발언된 발화를 인식하고, 이에 대한 응답으로, 발화의 오디오 표현과 상기 발화가 "John Doe"에 의해 발언되었다는 표시를 모두 포함하는 쿼리를 서버에 제공할 것이다.
- [0005] 이 조합을 통해 가상 어시스턴트 소프트웨어는 John Doe의 개인 정보에 액세스하여 쿼리에 대한 응답을 제공할 수 있다. 이 액세스를 용이하게 하기 위해 쿼리 재작성(re-write)이 발생할 수도 있다 (예: "오늘 일정에 있는 내용"의 초기 쿼리를 [[식별된 사용자]의 [날짜]에 대한 일정에 기록"). 동시에, 시스템은 단순히 한 명 이상의 가능한 화자의 발화, 발화 및 표시("id") 중 하나, 또는 장치 ID와 같은 일부 다른 표시(indication) 및 발화를 서버에 제공함으로써 인식되지 않는 당사자(예를 들어, 게스트 사용자)에 의한 쿼리 프로세싱을 허용하도록 구성된다.
- [0006] 서버는 쿼리(질의)에 대한 응답을 생성하기 위해 특정 사용자의 표시 또는 그러한 표시의 부재를 사용하여 쿼리에 대한 응답을 생성할 수 있다. 예를 들어, 서버는 발화는 John Doe에 의해 발언되었다는 표시의 수신에 기초하여, 쿼리에 응답하여 적절한 개인 정보를 제공할 권한이 있다고 결정할 수 있다. 하루 일정에 대한 요청의 예에서, 이는 서버가 John Doe에 대한 약속의 목록 또는 요약을 발성(vocalization)하기 위해 음성 지원 장치에 제공할 수 있음을 의미한다. 서버가 쿼리가 특정 사용자의 식별(식별정보)(identification)을 나타내지 않는다고 결정했지만 쿼리가 개인 정보를 찾고 있는 경우, 서버는 사용자가 음성 지원 장치에 알려지지 않아서 응답을 제공할 수 없음을 나타낼 수 있다. 또 다른 예에서, 서버는 발화에 대응하는 쿼리(예를 들어, "몇시입니까")가 특정 사용자를 나타내지 않더라도, 쿼리는 개인 정보와 관련이 없으며 응답(예: "현재 오전 11시 입니다")이 음성 지원 장치에 제공될 수 있다고 결정할 수 있다.
- [0007] 음성 지원 장치는 사용자가 쿼리를 제공함에 따라 사용자의 음성 인식을 적응시킬 수 있다. 예를 들어, 사용자를 인식하기 위해, 음성 지원 장치는 등록 프로세스 동안 말하는 사용자의 샘플과 쿼리를 제공하는 사용자의 최근 샘플의 조합을 사용할 수 있다. 음성 지원 장치는 그 장치가 각각의 사용자가 쿼리를 발언했다고 결정한 후에 쿼리로부터 각각의 사용자에게 대한 새로운 샘플을 생성한 다음, 그 새로운 샘플을 사용하고 추후 인식을 위해 더 오래된 샘플의 사용을 중지할 수 있다. 추가적으로 또는 대안적으로, 샘플은 특정 사용자에게 의해 사용되는 다수의 상이한 음성 지원 장치에 걸쳐 공유될 수 있어서, 각 장치는 특정 사용자가 쿼리를 제공할 때 유사하게 인식할 수 있다.
- [0008] 일부 양태에서, 본 명세서에서 설명된 요지는 동작을 포함할 수 있는 방법으로 구현될 수 있으며, 상기 동작은, 알려진 장치 사용자에게 대응하는 인증 토큰을 저장하는 동작; 화자로부터 발화를 수신하는 동작, 알려진 사용자

들 중 특정 알려진 사용자가 발언한 것으로 상기 발화를 분류하는 동작; 상기 발화의 표현(representation) 및 특정 알려진 사용자의 인증 토큰을 사용하여 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 동작을 포함한다.

[0009] 일부 양태들에서, 상기 발화의 표현 및 특정 알려진 사용자의 인증 토큰을 사용하여 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 동작은, 인증 토큰 및 발화가 특정 알려진 사용자가 발언한 것으로 분류되었음을 나타내는 플래그를 포함하는 쿼리를 제공하는 동작을 포함한다. 특정 양태들에서, 상기 발화의 표현 및 특정 알려진 사용자의 인증 토큰을 사용하여 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 동작은, 알려진 특정 사용자의 인증 토큰을 포함하는 쿼리를 제공하는 동작을 포함하며, 쿼리에서 특정 알려진 사용자의 인증 토큰을 포함하는 것은 발화가 특정 알려진 사용자에 의해 발언된 것으로 분류되었음을 나타낸다. 일부 구현들에서, 알려진 장치 사용자에 대응하는 인증 토큰을 저장하는 동작은 알려진 장치 사용자 각각에 대한 인증 토큰을 저장하는 동작을 포함하며, 여기서 각각의 인증 토큰은 알려진 사용자들 각각에 대응한다. 일부 양태들에서, 알려진 장치 사용자에 대응하는 인증 토큰들을 저장하는 것은 음성 지원 장치의 알려진 사용자들로서 등록하는 사용자들에 대한 응답이다.

[0010] 특정 양태들에서, 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 발화를 분류하는 것은 상기 발화가 특정 알려진 사용자에 대응하는 음성과 매칭하는 것으로 결정하는 것을 포함한다. 일부 구현들에서, 상기 발화가 특정 알려진 사용자에 대응하는 음성과 매칭하는 것으로 결정하는 것은, 발화가 미리 결정된 구문을 포함한다고 결정하는 것; 발화가 미리 결정된 구문을 포함한다는 결정에 응답하여, 미리 결정된 구문의 발화가 특정 알려진 사용자에 의한 미리 결정된 구문의 사전 발화와 매칭(일치)한다고 결정하는 것을 포함한다. 일부 양태들에서, 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 발화를 분류하는 것은 화자에 대응하는 시각적 정보(시각 정보)가 특정 알려진 사용자에 대응하는 시각적 정보와 매칭한다고 결정하는 것을 포함한다.

[0011] 특정 양태에서, 화자에 대응하는 시각적 정보(시각 정보)가 특정 알려진 사용자에 대응하는 시각적 정보와 매칭한다고 결정하는 것은 화자의 지문, 망막 스캔, 얼굴 또는 자세 중 하나 이상을 설명하는 화자 정보를 수신하는 것과, 화자 정보가 특정 알려진 사용자에 대응하는 시각적 정보와 매칭하는 것으로 결정하는 것을 포함한다. 일부 양태들에서, 발화의 표현 및 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 것은 쿼리를 서버에 제공하는 것을 포함한다. 일부 구현예들에서, 동작들은 서버로부터 쿼리에 대한 응답을 수신하고 그 응답을 화자에 제공하는 것을 포함한다.

[0012] 일부 양태에서, 본 명세서에서 설명된 요지는 동작을 포함할 수 있는 방법으로 구현될 수 있으며, 상기 동작은, 알려진 장치 사용자에 대응하는 인증 토큰을 저장하는 동작; 화자로부터 발화를 수신하는 동작, 알려진 사용자들 중 어느 누구도 발언하지 않은 것으로 상기 발화를 분류하는 동작, 발화의 표현을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 분류되었음을 나타내지 않는 쿼리를 제공하는 동작을 포함한다. 특정 양태들에서, 발화의 표현을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 분류되었음을 나타내지 않는 쿼리를 제공하는 동작은 인증 토큰을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자에 의해 발언된 것으로 분류됨을 나타내는 플래그를 포함하지 않는 쿼리를 제공하는 동작을 포함한다. 일부 구현예에서, 발화의 표현을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 분류되었음을 나타내지 않는 쿼리를 제공하는 동작은 알려진 사용자들의 인증 토큰을 포함하지 않는 쿼리를 제공하는 동작을 포함한다.

[0013] 일부 양태에서, 본 명세서에서 설명된 요지는 다음의 동작을 포함할 수 있는 방법으로 구현될 수 있으며, 상기 동작은 쿼리가 특정 알려진 사용자를 화자로 나타내는지 결정하는 동작; 쿼리가 특정 알려진 사용자를 화자로서 나타내는지 결정하는 것에 응답하여, 쿼리가 알려진 사용자의 것이 아니라고 결정하는 동작; 쿼리가 알려진 사용자의 것이 아니라고 결정한 것에 응답하여, 쿼리가 비-개인적(non-personal)인지 여부를 결정하는 동작; 그리고 쿼리가 비-개인적이라는 결정에 응답하여, 쿼리에 응답을 제공하려고 시도하는 동작을 포함한다. 특정 양태들에서, 쿼리가 특정 알려진 사용자를 화자로서 나타내는지 결정하는 동작은 발화를 나타내는 플래그를 포함하는 쿼리가 특정 알려진 사용자가 발언한 것으로 분류되는지를 결정하는 동작을 포함한다. 일부 구현예들에서, 쿼리가 특정 알려진 사용자를 화자로 나타내는지 결정하는 동작은 알려진 특정 사용자의 인증 토큰을 포함하는 쿼리를 결정하는 동작을 포함한다. 일부 양태들에서, 쿼리가 알려진 사용자의 것이 아니라고 결정한 것에 응답하여, 쿼리가 비-개인적인 것인지를 결정하는 동작은 쿼리가 알려진 사용자의 것이 아니라고 결정한 것에 응답하여, 쿼리에 대한 응답(답변)이 개인 정보에 의존하는지 여부를 결정하는 동작을 포함한다.

[0014] 일부 양태에서, 본 명세서에서 설명된 요지는 다음의 동작을 포함할 수 있는 방법으로 구현될 수 있으며, 상기

동작은, 발화가 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 적어도 기초하여 핫워드를 말하고 있는 특정 사용자를 포함하는 것을 결정하는 동작; 발화가 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 적어도 기초하여 핫워드를 말하고 있는 특정 사용자를 포함한다는 결정에 응답하여, 발화의 적어도 일부 (부분)를 새로운 샘플로서 저장하는 동작; 발화를 말하고 있는 특정 사용자의 제2 세트의 샘플들을 획득하는 동작 -상기 제2 세트의 샘플들은 새로운 샘플을 포함하며, 제1 세트의 샘플들의 모든 샘플보다 적음-; 적어도 핫워드를 말하고 있는 사용자의 제2 세트의 샘플들에 기초하여 제2 발화가 핫워드를 말하고 있는 특정 사용자를 포함하는 것으로 결정하는 동작을 포함한다.

[0015] 특정 양태에서, 발화를 말하고 있는 특정 사용자의 제2 세트의 샘플들을 획득하는 동작 -제2 세트의 샘플들은 새로운 샘플을 포함하고 제1 세트의 샘플들 내의 모든 샘플보다 적은 샘플을 포함함 -은 미리 결정된 수의 최근 저장된 샘플들을 제2 세트의 샘플들로서 선택하는 동작을 포함한다. 일부 양태에서, 발화를 말하고 있는 특정 사용자의 제2 세트의 샘플들을 획득하는 동작 -제2 세트의 샘플들은 새로운 샘플을 포함하고 그리고 제1 세트의 샘플들 내의 모든 샘플보다 적은 샘플을 포함함 -은 미리 결정된 개수의 가장 최근에 저장된 샘플들과 레퍼런스 샘플들의 세트를 모두 선택하여 제2 세트의 샘플들로서 함께 결합하는 동작을 포함한다. 일부 구현들에서, 레퍼런스 샘플들은 특정 사용자에 대한 등록 프로세스로부터의 샘플들을 포함하고 그리고 가장 최근에 저장된 샘플들은 특정 사용자가 발언한 쿼리로부터의 샘플들을 포함한다.

[0016] 특정 양태들에서, 동작들은 제2 세트의 샘플들을 획득하는 것에 응답하여, 제1 세트의 샘플들에서 샘플을 삭제하지만 제2 세트의 샘플들에서는 샘플을 삭제하지 않는 동작을 포함한다. 일부 양태들에서, 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 기초하여 발화가 핫워드를 말하고 있는 특정 사용자를 포함한다고 결정하는 동작은, 제1 세트의 샘플들을 사용하여 핫워드 검출 모델을 생성하는 동작; 핫워드 검출 모델에 발화를 입력하는 동작; 핫워드 검출 모델이 발화를 핫워드를 말하고 있는 특정 사용자를 포함하는 것으로 분류했다고 결정하는 동작을 포함한다. 일부 구현예들에서, 핫워드를 말하고 있는 사용자의 제2 세트의 샘플들에 기초하여 제2 발화가 핫워드를 말하고 있는 특정 사용자를 포함한다고 결정하는 동작은, 상기 제2 세트의 샘플들을 사용하여 제2 핫워드 검출 모델을 생성하는 동작; 제2 발화를 제2 핫워드 검출 모델에 입력하는 동작; 제2 핫워드 검출 모델이 핫워드를 말하고 있는 특정 사용자를 포함하는 것으로 제2 발화를 분류했다고 결정하는 동작을 포함한다.

[0017] 특정 양태에서, 동작들은 서버에서 제2 새로운 샘플을 수신하는 동작, 서버로부터의 제2 새로운 샘플을 포함하고 그리고 제2 세트의 샘플들 내의 모든 샘플보다 적은 제3 세트의 샘플들에 적어도 기초하여, 핫워드를 말하고 있는 특정 사용자를 제3 발화가 포함하는지를 결정하는 동작을 포함한다. 일부 양태에서, 동작들은 서버로부터, 제3 세트의 샘플들에서 샘플들의 표시를 수신하는 동작, 로컬로 저장되지 않은 제3 세트의 샘플들에 있는 샘플들을 결정하는 동작, 로컬로 저장되지 않은 제3 세트의 샘플들 내의 샘플들에 대한 요청을 서버에 제공하는 동작, 상기 요청에 응답하여 서버로부터 로컬로 저장되지 않은 샘플들을 수신하는 동작을 포함한다.

[0018] 일부 구현예들에서, 동작들은, 상기 음성 지원 장치가 상기 특정 사용자가 핫워드를 말하는지 여부를 검출할 수 있도록 상기 음성 지원 장치에 제1 세트의 샘플들을 제공하는 동작을 포함하며, 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 기초하여 발화가 핫워드를 말하고 있는 특정 사용자를 포함하는지를 결정하는 동작은, 상기 음성 지원 장치가 상기 특정 사용자가 핫워드를 말했음을 검출했다는 표시를 수신하는 동작을 포함한다. 특정 양태들에서, 동작은, 상기 제1 세트의 샘플들을 사용하여 핫워드 검출 모델을 생성하고, 상기 핫워드 검출 모델을 음성 지원 장치에 제공하여 상기 음성 지원 장치가 상기 특정 사용자가 상기 핫워드를 말했는지 여부를 검출할 수 있도록 하는 동작을 포함하며, 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 적어도 기초하여 발화가 핫워드를 말하고 있는 특정 사용자를 포함하는 것으로 결정하는 동작은 상기 음성 지원 장치가 상기 특정 사용자가 핫워드를 말했음을 검출했다는 표시를 수신하는 동작을 포함한다.

[0019] 일부 구현예들에서, 동작은, 특정 사용자가 핫워드를 말했는지 여부를 검출하기 위해 현재 세트의 샘플들에 대한 요청을 음성 지원 장치로부터 수신하는 동작; 음성 지원 장치에 의해 로컬적으로 저장되지 않은 현재 세트의 샘플들의 샘플들을 결정하는 동작; 상기 음성 지원 장치에, 상기 음성 지원 장치에 의해 로컬적으로 저장되지 않은 현재 세트의 샘플들의 샘플들 및 상기 현재 세트의 샘플들 내의 샘플들의 표시를 제공하는 동작을 포함한다.

[0020] 이 양태 및 다른 양태의 다른 구현 예는 컴퓨터 저장 장치상에 인코딩된 방법의 동작을 수행하도록 구성된 시스템, 장치 및 컴퓨터 프로그램을 포함한다. 하나 이상의 컴퓨터로 구성된 시스템은 소프트웨어, 펌웨어, 하드웨어 또는 시스템에 설치된 이들의 조합으로 작동하여 시스템이 동작들을 수행하도록 구성될 수 있다. 하나 이상

의 컴퓨터 프로그램은 데이터 처리 장치에 의해 실행될 때 그 장치로 하여금 동작을 수행하게 하는 명령(명령어)을 가짐으로써 구성될 수 있다.

[0021] 본 명세서에 기술된 요지의 특정 실시 예는 다음 장점 중 하나 이상을 실현하도록 구현될 수 있다. 예를 들어, 음성 지원 장치가 발화의 인증 토큰 화자를 포함하는 쿼리를 서버에 제공하게 함으로써, 시스템은 음성 지원 장치가 다수의 상이한 사용자에게 의해 공유되고 발화의 화자에게 개인화된 응답을 제공할 수 있게 할 수 있다. 다른 예에서, 음성 지원 장치에 저장된 사용자에 대한 다수 또는 모든 인증 토큰을 포함하는 쿼리를 갖고, 어떤 인증 토큰이 화자에 대응하는지 또는 인증 토큰이 화자에 대응하지 않는지를 나타냄으로써, 시스템은 서비스가 상기 공유된 음성 지원 장치의 사용자들 사이에서 공유될 수 있게 한다. 예를 들어, 쿼리를 프로세싱하는 서버가 음악 스트리밍 서비스를 사용하도록 허용된 특정 사용자에 대한 인증 토큰을 포함하는 토큰 풀(pool)이 포함된 쿼리를 수신한 경우, 상기 쿼리가 특정 사용자가 발화를 말한 것으로 나타내지 않더라도, 서버는 특정 사용자에 대한 인증 토큰이 존재하는 경우, 쿼리에 응답하여 음악 스트리밍 서비스가 여전히 사용되도록 허용할 수 있다. 따라서, 시스템은 예를 들어 게스트 사용자들에 의한 음성 지원 장치의 일반적인 사용을 여전히 허용하면서 특정 사용자의 개인 정보에 대한 액세스를 보안을 할 수 있다. 따라서, 시스템은 데이터 보안과 관련된 문제를 해결할 수 있다. 시스템은 다수의 상이한 사용자에게 의해 사용되는 장치에서 개인화된 사용자 상호 작용을 제공하는 방법과 관련된 문제를 추가로 또는 대안적으로 해결할 수 있다.

[0022] 다른 예에서, 시스템이 발화를 말하는 사용자들의 새로운 샘플들에 기초하여 상이한 세트의 샘플들을 획득함으로써, 시스템은 음성, 악센트, 환경, 또는 음성 지원 장치에 의해 수신된 오디오의 다른 요소가 변경될 때에도 특정 알려진 사용자가 발화를 말할 때 검출의 정확성을 보장할 수 있다. 또 다른 예에서, 시스템이 다수의 상이한 음성 지원 장치들 사이에서 샘플들을 전송하게 함으로써, 시스템은 사용자가 말을 할 필요가 없이, 각각의 음성 지원 장치에 대한 레퍼런스 샘플을 제공함으로써 시간을 절약할 수 있게 하고, 그리고 음성 지원 장치들 각각이 핫워드를 말하는 알려진 사용자를 유사하게 검출하도록 보장할 수 있다. 따라서, 사용자가 제1 등록 후에 말함으로써 레퍼런스 샘플을 제공하기 위해 필요한 프로세싱이 절약될 수 있다. 예를 들어, 시스템은 특정 발화가 알려진 사용자의 하나의 음성 지원 장치에서 말하는 것으로 인식되면, 알려진 사용자의 모든 음성 지원 장치들은 알려진 사용자가 말한 것으로 상기 특정 발화를 유사하게 인식할 것이다. 따라서, 알려진 사용자에게 대한 일관성 및 예측성이 증가될 수 있다.

[0023] 본 명세서에서 설명된 요지의 하나 이상의 구현예들의 세부 사항은 첨부 도면 및 이하의 설명에서 설명된다. 본 요지의 다른 특징, 양태 및 잠재적 장점은 상세한 설명, 도면 및 청구 범위로부터 명백해질 것이다.

**도면의 간단한 설명**

[0024] 도 1a-1c는 음성 지원 장치와의 예시적인 상호 작용을 나타내는 블록도이다.  
 도 2는 알려진 사용자로부터 발화를 처리하기 위한 프로세스의 예를 나타내는 흐름도이다.  
 도 3은 게스트 사용자로부터 발화를 처리하기 위한 프로세스의 예를 나타내는 흐름도이다.  
 도 4는 음성 지원 장치로부터의 쿼리를 처리하기 위한 프로세스의 예를 나타내는 흐름도이다.  
 도 5는 음성 지원 장치로부터의 쿼리를 처리하기 위한 프로세스의 다른 예를 나타내는 흐름도이다.  
 도 6은 화자 식별에 사용되는 샘플 세트를 업데이트하는 예를 나타내는 블록도이다.  
 도 7은 화자 식별에 사용되는 샘플 세트를 업데이트하기 위한 프로세스의 예를 나타내는 흐름도이다.  
 도 8은 컴퓨팅 장치의 예의 도면이다.  
 다양한 도면에서 유사한 참조 번호 및 명칭은 유사한 요소를 나타낸다.

**발명을 실시하기 위한 구체적인 내용**

[0025] 도 1a 내지 도 1c는 시스템(100)에서의 상이한 예시적인 상호 작용을 나타내는 블록도이다. 시스템(100)은 음성 지원 장치(speech-enabled device)(125) 및 서버(155)를 포함한다. 음성 지원 장치(125)와 서버(155)는 함께 작동하여 사용자(110)에 의한 발화(120)에 대한 응답을 제공할 수 있다. 예를 들어, 음성 지원 장치(125)는 음성 지원 장치(125)에 근접한 사용자에게 의해 발언된 발화에 기초하여 서버에 쿼리를 제공할 수 있다. 서버(155)는 발화에 대한 응답을 음성 지원 장치(125)에 제공할 수 있다.

- [0026] 일부 구현예들에서, 음성 지원 장치(125)는 사용자가 시스템(100)으로부터 응답을 시작하기 위해 말하는 핫워드라고 하는 미리 결정된 구문을 포함하는 발화의 검출에 응답하여 서버(155)에 쿼리를 제공할 수 있다. 예를 들어, 핫워드는 "오케이 컴퓨터(OK Computer)" 또는 다른 구문일 수 있다.
- [0027] 다수의(복수의) 사용자로부터의 발화를 처리하기 위해, 음성 지원 장치 (125)는 특정 알려진 사용자에게 의해 발언된 것으로 발화를 분류하고, 발화를 발언한 것으로 분류된 특정 알려진 사용자를 나타내는 쿼리를 서버에 제공할 수 있다. 알려진 사용자는 시스템(100)의 사용자로서 등록된 사용자 일 수 있고, 게스트 사용자는 시스템(100)의 사용자로서 등록되지 않은 사용자일 수 있다. 예를 들어, "아빠(Dad)"는 음성 지원 장치(125)의 알려진 사용자로서 등록될 수 있고, 음성 지원 장치(125)는 발화가 알려진 사용자 "Dad"에 의해 발언되는지를 나중에 분류할 수 있다. 서버(155)는 쿼리에 대한 응답을 생성하기 위해, 쿼리에서 특정 알려진 사용자의 표시(indication) 또는 쿼리에 그러한 표시의 부재(lack)를 사용할 수 있다.
- [0028] 예를 들어, 도 1a는 알려진 사용자 "아빠"에 의해 발언된 "오케이 컴퓨터, 내 일정이 어떻게 되지"라는 발화를 설명한다. 음성 지원 장치(125)는 서버(155)에 쿼리를 제공하는데, 여기서 쿼리는 발화의 표현 및 발화가 알려진 사용자 "아빠(Dad)"에 의해 발언된 것으로 분류되었다는 표시(indication)를 포함한다. 서버(155)는 알려진 사용자 "Dad"의 표시에 기초하여 "아빠, 오늘 오후 1시에 존과 1회 미팅이 있어요"라는 응답을 생성한다.
- [0029] 다른 예에서, 도 1b는 게스트 사용자에게 의해 발언된 발화 "오케이 컴퓨터, 내 일정이 어떻게 되지"를 나타낸다. 음성 지원 장치(125)는 서버(155)에 쿼리를 제공하는데, 여기서 쿼리는 발화의 표현을 포함하고 발화가 음성 지원 장치(125)의 임의의 알려진 사용자에게 의해 발언된 것으로 분류되었다는 표시를 포함하지 않는다. 그 후, 서버(155)는 발화가 음성 지원 장치(125)의 임의의 알려진 사용자에게 의해 발언된 것으로 분류되었다는 표시의 부재에 기초하여 "일정에 액세스할 수 없지만, 오늘은 전국 아이스크림의 날이라는 것을 알고 있었나요?"라는 응답을 생성한다.
- [0030] 또 다른 예에서, 도 1c는 게스트 사용자에게 의해 발언된 발화 "오케이 컴퓨터, 몇시지?"를 나타낸다. 음성 지원 장치(125)는 서버(155)에 쿼리를 제공하는데, 여기서 쿼리는 발화의 표현을 포함하고 발화가 음성 지원 장치(125)의 임의의 알려진 사용자에게 의해 발언된 것으로 분류되었다는 표시를 포함하지 않는다. 그리고 나서 서버(155)는 "오전 11시입니다"라는 응답을 생성하고, 그리고 알려진 사용자가 말하지 않았지만, 쿼리는 개인 정보와 관련이 없다는 결정에 기초하여 음성 지원 장치(125)에 상기 응답을 제공한다.
- [0031] 구체적으로, 음성 지원 장치(125)는 하나 이상의 마이크로폰 및 하나 이상의 스피커를 포함할 수 있다. 음성 지원 장치(125)는 하나 이상의 마이크로폰을 사용하여 발화를 수신하고 그리고 하나 이상의 스피커를 통해 발화에 대한 가청 응답을 출력할 수 있다. 음성 지원 장치(125)는 각각의 알려진 사용자에게 대응하는 인증 토큰을 저장할 수 있다. 예를 들어, 음성 지원 장치는 제1 알려진 사용자 "엄마"에 대한 제1 인증 토큰(132), 제2 알려진 사용자 "아빠"에 대한 제2 인증 토큰(134), 제3 알려진 사용자 "매트"에 대한 제3 인증 토큰(136), 제4 알려진 사용자 "케이티"를 위한 제4 인증 토큰(138)을 저장할 수 있다.
- [0032] 사용자의 인증 토큰은 시스템(100)에서 사용자의 계정에 대한 로그인 역할을 하는 데이터 구조일 수 있다. 예를 들어, 알려진 사용자 "아빠"(134)에 대한 인증 토큰을 포함하고 "아빠"가 화자로 분류됨을 나타내는 쿼리는 "아빠"의 계정에 있는 정보에 액세스할 수 있다. 인증 토큰은 발화에 응답하여 수행될 수 있는 동작을 나타낼 수 있다. 예를 들어, 알려진 사용자 "아빠"의 계정은 "Provider X"에서 음악을 재생할 수 있는 권한을 가질 수 있으므로, "아빠"(134)에 대한 인증 토큰을 포함하는 "Provider X"에서 음악을 재생하기 위한 쿼리로 인해 "Provider X"에서 음악이 재생될 수 있다. 다른 예에서, 현재 화자뿐만 아니라 "Provider X"에서 음악을 재생할 수 있는 권한이 있는 계정을 가진 알려진 사용자에게 대한 인증 토큰을 포함하지 않는 "Provider X"의 음악 재생 쿼리로 인해 음악이 재생되지 않을 수 있다.
- [0033] 일부 구현예들에서, 음성 지원 장치(125)는 음성 지원 장치의 알려진 사용자로서 등록한 사용자에게 응답하여 알려진 사용자에게 대한 인증 토큰을 저장할 수 있다. 예를 들어, "엄마", "아빠", "매트" 및 "케이티" 사용자는 음성 지원 장치(125)의 알려진 사용자로서 명시적으로 등록될 수 있다. 등록 프로세스에서, 음성 지원 장치(125)는 서버(155)로부터 알려진 사용자에게 대한 인증 토큰을 수신할 수 있다. 예를 들어, "아빠"가 음성 지원 장치(125)의 알려진 사용자로서 등록될 때, 서버(155)는 "아빠"(134)에 대한 인증 토큰을 음성 지원 장치(125)에 제공할 수 있다. 등록 프로세스에서, 음성 지원 장치(125)는 또한 화자를 알려진 사용자로서 식별하는데 사용될 수 있는 컨텍스트 정보를 저장할 수 있다. 예를 들어, "아빠"가 음성 지원 장치(125)의 알려진 사용자로서 등록될 때, 음성 지원 장치(125)는 "아빠"가 핫워드를 여러 번 말하도록 요청한 다음, 핫워드의 발화에 대응하는 화자

식별 특징을 로컬적으로(locally) 저장할 수 있다.

- [0034] 일부 구현들에서, 사용자는 모바일 컴퓨팅 장치가 로컬 무선 연결을 통해 음성 지원 장치(125)와 통신하는 모바일 컴퓨팅 장치상의 컴패니언(companion) 애플리케이션을 통해 알려진 사용자로서 등록할 수 있다. 예를 들어, 사용자는 스마트 폰에서 컴패니언 애플리케이션을 통해 사용자의 계정에 로그인한 다음, 컴패니언 애플리케이션에서 사용자가 음성 지원 장치(125)의 알려진 사용자로서 등록하기를 원한다고 지시한 다음, 핫워드를 모바일 컴퓨팅 장치에 여러 번 말할 수 있다. 컴패니언 애플리케이션은 사용자에게 대한 인증(authentication)이 사용자가 핫워드를 말하는 것에 기초하여 화자 식별 특징과 함께 음성 지원 장치(125) 상에 저장되게 할 수 있다.
- [0035] 일부 구현예들에서, 음성 지원 장치(125)는 쿼리에서 특정 알려진 사용자에게 대한 인증 토큰에 대한 플래그를 포함하는 것에 기초하여 쿼리에서 특정 알려진 사용자의 표시를 포함할 수 있다. 예를 들어, 음성 지원 장치(125)는 쿼리로 음성 지원 장치(125)에 저장된 모든 인증 토큰(본 명세서에서 "풀(pool)"이라고 함)을 포함할 수 있고, 음성 지원 장치(125)가 인증 토큰에 대응하는 식별된 알려진 사용자에게 의해 발언된 발화로서 분류하는 인증 토큰에 대한 플래그를 설정할 수 있다.
- [0036] 이러한 구현예들에서, 인증 토큰들의 풀은 응답이 알려진 사용자의 개인 정보에 의존하지 않는 쿼리를 처리하기 위해 서버(155)에 의해 사용될 수 있다. 예를 들어, 쿼리는 "X 서비스에서 음악 재생"일 수 있고, 그리고 서버(155)는 "X 서비스로부터 음악 재생"에 대한 권리를 갖는 알려진 사용자와 연관된 풀로부터 인증 토큰을 식별하기 위해 쿼리에서 인증 토큰을 검사할 수 있다. 서버(155)가 "X 서비스로부터 음악 재생"에 대한 권리를 갖는 적어도 하나의 인증 토큰을 식별하면, 서버(155)는 "X 서비스"로부터의 음악을 재생할 수 있다. 이 예에서, 쿼리에 풀로서 모든 인증 토큰을 포함시키는 것은 게스트 사용자를 포함하는 음성 지원 장치(125)의 임의의 사용자가 음성 지원 장치(125)로 하여금 "X 서비스"로부터 음악을 재생할 수 있게 할 수 있다.
- [0037] 일부 구현예들에서, 음성 지원 장치(125)는 쿼리에 특정 알려진 사용자에게 대한 인증 토큰을 포함함으로써 (즉, 플래그를 포함하지 않음) 쿼리에 특정 알려진 사용자의 표시를 포함할 수 있다. 예를 들어, 일부 구현예들에서, 음성 지원 장치(125)는, 음성 지원 장치(125)가 인증 토큰에 대응하는 알려진 사용자에게 의해 발언된 것으로 발화를 분류할 때 쿼리에 인증 토큰만을 포함한다. 이러한 구현예들에서, 알려진 사용자와 일치하지 않는 게스트 사용자의 발화가 있는 경우, 음성 지원 장치(125)는 알려진 사용자들 중 어느 누구도 말하지 않은 것으로 발화를 분류하고 인증 토큰을 포함하지 않는 쿼리를 서버(155)에 제공한다.
- [0038] 일부 구현예들에서, 음성 지원 장치(125)는, 음성 지원 장치(125)가 가질 수 있는 권한을 나타내는 장치(디바이스) 인증 토큰을 저장할 수 있고, 서버(155)에 대한 쿼리들에 장치 인증 토큰을 포함할 수 있다. 예를 들어, 음성 지원 장치(125)가 "Provider X"로부터 음악을 재생할 수 있는 권한이 있는 알려진 사용자에게 대한 인증 토큰을 저장하면, 음성 지원 장치(125)는 음성 지원 장치(125)가 "Provider(공급자) X"로부터 음악을 재생할 수 있는 권한이 있음을 나타내는 장치 인증 토큰을 저장할 수 있고 그리고 모든 사용자가 "Provider(공급자) X"로부터 음악을 재생할 수 있도록(즉, 특정 사용자 토큰이 필요하지 않음) 모든 쿼리에 장치 인증 토큰을 포함시킨다. 다른 예에서, 음성 지원 장치(125)가 "공급자 X"로부터 음악을 재생할 수 있는 권한이 있는 알려진 사용자에게 대한 인증 토큰을 저장하지 않는 경우, 음성 지원 장치(125)는, 음성 지원 장치(125)가 "공급자 X"로부터 음악을 재생할 권한이 있음을 나타내지 않는 장치 인증 토큰을 저장할 수 있다.
- [0039] 일부 구현예들에서, 음성 지원 장치(125)는 컨텍스트 정보에 기초하여 특정 사용자에게 의해 발언된 것으로 발화를 분류할 수 있다. 컨텍스트 정보는 오디오, 시각 또는 다른 정보 중에서 하나 이상을 포함할 수 있다. 오디오 정보와 관련하여, 음성 지원 장치(125)는 알려진 사용자의 하나 이상의 발화의 화자 식별 특징(예를 들어, 집합적으로 특징 벡터를 형성할 수 있는 MFCC(Mel-Frequency Cepstral Coefficient) 특징)에 기초하여 발화를 분류할 수 있다. 예를 들어, 음성 지원 장치(125)는 "오케이 컴퓨터"를 말하고 있는 알려진 사용자 각각에 대한 화자 식별 특징을 저장할 수 있다. "오케이 컴퓨터"라고 말하고 있는 알려진 사용자 "아빠"의 저장된 화자 식별 특징과 충분히 매칭하는 현재 수신된 발화의 화자 식별 특징에 응답하여, 음성 지원 장치(125)는 알려진 사용자 "아빠"에 의해 발언된 것으로 발화를 분류할 수 있다.
- [0040] 다른 예에서, 음성 지원 장치(125)는 발화의 전체 오디오에 기초하여 발화를 분류할 수 있다. 예를 들어, 음성 지원 장치(125)는 전체 수신된 발화에서의 음성이 알려진 사용자 "아빠"에 대응하는 음성과 매칭하는지 여부를 결정할 수 있다.
- [0041] 시각 정보와 관련하여, 음성 지원 장치(125)는 화자의 적어도 일부(부분)의 하나 이상의 이미지를 수신할 수 있고 그리고 하나 이상의 이미지를 기반으로 화자를 인식할 수 있다. 예를 들어, 음성 지원 장치(125)는 카메라를

포함할 수 있고 그리고 카메라의 뷰 내의 화자가 음성 지원 장치(125)가 알려진 사용자 "아빠"에 대응하는 얼굴과 매칭하는 것으로 분류한 얼굴을 갖는지를 결정할 수 있다. 다른 예들에서, 음성 지원 장치(125)는 화자의 지문, 망막 스캔, 얼굴 인식, 자세, 다른 장치의 공존, 또는 다른 장치 또는 소프트웨어 요소로부터의 신원 확인 중 하나 이상을 매칭시키려고 시도할 수 있다.

[0042] 서버(155)는 음성 지원 장치(125)로부터 쿼리를 수신하고, 음성 지원 장치(125)에 쿼리에 대한 응답을 제공할 수 있다. 예를 들어, 서버(155)는 "오케이 컴퓨터, 내 일정은?"이라는 쿼리를 수신하고, 이에 대한 응답으로 "아빠, 오늘 오후 1시에 존과 회의가 한 번 있습니다."라는 응답을 제공할 수 있다. 서버(155)는 도 1a 내지 1c에 도시된 바와 같이, 쿼리가 알려진 사용자의 표시를 포함하는지 여부 및 쿼리가 개인 정보와 관련된 질문을 포함하는지 여부에 기초하여 쿼리에 대한 응답을 제공할 수 있다.

[0043] 구체적으로, 서버(155)는 발화가 음성 지원 장치(125)에 의해 알려진 사용자가 말한 것으로 분류되었음을 쿼리가 나타내는지를 초기에 결정할 수 있다. 예를 들어, 쿼리가 음성 지원 장치(125)에 저장된 모든 인증 토큰의 풀을 포함하는 경우, 서버(155)는 쿼리가 "아빠"의 인증 토큰에 대해 "참(True)" 값을 갖는 "베스트-화자(best-speaker)"의 플래그를 포함한다고 결정할 수 있다. 다른 예에서, 쿼리가 발화를 말하는 것으로 분류된 알려진 사용자에게 대한 인증 토큰만을 포함하는 경우, 서버(155)는 쿼리가 "아빠"의 인증 토큰을 포함한다고 결정할 수 있다.

[0044] 서버(155)는 발화가 알려진 사용자에게 의해 발언된 것으로 음성 지원 장치(125)에 의해 분류되었음을 쿼리가 나타내는 것으로 결정하면, 서버(155)는 알려진 사용자와 관련하여 쿼리를 프로세싱할 수 있다. 예를 들어, 발화 "오케이 컴퓨터, 내 일정은?"의 오디오 표현을 포함하고 알려진 사용자 "아빠"를 나타내는 쿼리의 경우, 서버(155)는 오디오 표현으로부터 "오케이 컴퓨터, 내 일정은?"의 전사(transcription)를 생성할 수 있고, 그 전사에 기초하여 알려진 사용자 "아빠"의 일정(calendar)에 액세스하고, "아빠, 오늘 오후 1시에 존과 한 번의 회의가 있어요"라는 응답을 생성하고, 그 응답을 음성 지원 장치(125)에 제공할 수 있다. 선택적으로, 서버(155) (또는 이와 관련된 모듈)는 전사를 재기록하기 위해 알려진 화자 식별 및/또는 오디오 표현을 사용할 수 있다. 예를 들어, 화자가 "아빠"라는 것을 알면, 서버(155)는 "오케이 컴퓨터, 내 일정은?"으로부터 "오케이 컴퓨터, [아빠]의 일정은?"으로 쿼리를 재작성할 수 있다.

[0045] 서버(155)가, 발화가 알려진 사용자에게 의해 발언된 것으로 분류되었음을 쿼리가 나타내지 않는다고 결정하는 경우, 서버(155)는 쿼리가 개인 정보와 관련되어 있는지 여부를 결정할 수 있다. 예를 들어, 서버(155)는 질문에 대한 답변이 사용자에게 대한 개인 정보에 의존하기 때문에 (예를 들어, 질문에 응답하기 위해 보안 사용자 계정에 액세스해야 함), "오케이 컴퓨터, 내 일정은?"이라는 발화의 오디오 표현을 포함하는 쿼리가 개인 정보와 관련되어 있다고 결정할 수 있다. 다른 예에서, 서버(155)는 질문에 대한 답변이 사용자의 개인 정보에 의존하지 않기 때문에 발화 "오케이 컴퓨터, 몇시입니까?"의 음성 표현을 포함하는 쿼리가 개인 정보와 관련되지 않은 것으로 결정할 수 있다.

[0046] 서버(155)가, 발화가 알려진 사용자에게 의해 발언된 것으로 음성 지원 장치(125)에 의해 분류되지 않았다고 결정하고, 쿼리가 개인 정보와 관련되어 있다고 결정하면, 서버(155)는 쿼리에 대한 응답(답변)을 제공하지 않기로 결정할 수 있다 (및 선택적으로 추가 정보, 대화 또는 인식 부족을 정정하기 위한 가이드를 제공함). 예를 들어, 도 1b에 도시된 바와 같이, 게스트 사용자가 말한 발화 "오케이 컴퓨터, 내 일정은?"의 경우, 서버(155)는 "일정에 액세스할 수 없지만 오늘은 전국 아이스크림의 날이라는 것을 알고 있습니까?"라는 응답(또는 "일정에 액세스할 수 없지만, 원하실 경우 등록 절차를 안내할 수 있습니다.")을 제공한다.

[0047] 서버(155)가, 발화가 알려진 사용자에게 의해 발언된 것으로 음성 지원 장치(125)에 의해 분류됨을 나타내지 않는다고 결정하고, 쿼리가 개인 정보와 관련되지 않은 것으로 결정하면, 서버(155)는 쿼리에 대한 응답을 제공하기로 결정할 수 있다. 예를 들어, 도 1c에 도시된 바와 같이, 게스트 사용자에게 의해 발언된 "오케이 컴퓨터, 몇시입니까?"인 경우, 서버(155)는 "오전 11시입니다"라는 응답을 제공한다.

[0048] 음성 지원 장치(125) 및 서버(155)의 기능이 결합, 추가 분리, 분배 또는 교환될 수 있는 시스템(100)의 상이한 구성이 사용될 수 있다. 예를 들어, 서버(155)가 전사할 쿼리에 발화의 오디오 표현을 포함시키는 대신, 음성 지원 장치(125)는 발화를 전사하고 서버(155)에 대한 쿼리에 전사를 포함시킬 수 있다.

[0049] 도 2는 알려진 사용자로부터 발화를 처리하기 위한 프로세스(200)의 예를 나타내는 흐름도이다. 프로세스(200)의 동작은 도 1a-1c의 시스템(100)과 같은 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있다.

[0050] 프로세스(200)는 장치의 알려진 사용자에게 대응하는 인증 토큰을 저장하는 단계(205)를 포함한다. 예를 들어, 음

성 지원 장치(125)는 알려진 사용자 "엄마", "아빠", "매트" 및 "케이티" 각각에 대한 인증 토큰(132, 134, 136, 138)을 수신 및 저장할 수 있다. 인증 토큰들 각각은 알려진 사용자들 각각에 대응할 수 있다. 예를 들어, 제1 인증 토큰(132)은 제1 알려진 사용자 "엄마"에 대응할 수 있고, 제2 인증 토큰(134)은 제2 알려진 사용자 "아빠"에 대응할 수 있고, 제3 인증 토큰(136)은 제3 알려진 사용자 "매트"에 대응할 수 있고, 제4 인증 토큰(138)은 제4 알려진 사용자 "케이티"에 대응할 수 있다.

- [0051] 음성 지원 장치(125)는 사용자가 음성 지원 장치의 알려진 사용자로서 등록하는 것에 응답하여 인증 토큰을 저장할 수 있다. 예를 들어, "엄마"에 대한 인증 토큰은, 서버(155)가 "엄마"가 음성 지원 장치(125)를 사용할 것이라는 표시를 수신한 후 음성 지원 장치(125)가 서버(155)로부터 인증 토큰을 수신하는 것에 응답하여, 음성 지원 장치(125)에 의해 저장될 수 있다.
- [0052] 프로세스(200)는 화자로부터 발화를 수신하는 단계(210)를 포함한다. 예를 들어, 음성 지원 장치(125)는, 음성 지원 장치(125)의 마이크로폰을 통해 알려진 사용자 "아빠"로부터 "오케이 컴퓨터, 내 일정은?"이라는 발화를 수신할 수 있다.
- [0053] 프로세스(200)는 알려진 사용자들 중 특정 알려진 사용자가 말한 것으로 발화를 분류하는 단계(215)를 포함한다. 예를 들어, 음성 지원 장치(125)는 알려진 사용자 "아빠"에 의해 발언된 것으로 발화를 분류할 수 있다. 일부 구현예들에서, 알려진 사용자들 중 특정 알려진 사용자에 의해 발언된 것으로 발화를 분류하는 것은 발화가 특정 알려진 사용자에 대응하는 음성과 매칭한다고 결정하는 것을 포함한다. 예를 들어, 음성 지원 장치(125)는 발화의 화자 식별 특징이 "아빠"에 대해 이전에 저장된 화자 식별 특징과 매칭(일치)한다고 결정할 수 있다.
- [0054] 일부 구현예들에서, 발화가 특정 알려진 사용자에 대응하는 음성과 매칭한다고 결정하는 것은 발화가 미리 결정된 구문을 포함한다고 결정하고, 미리 결정된 구문의 발화를 결정하는 것에 응답하여, 미리 결정된 구문의 발화가 특정 알려진 사용자에 의한 미리 결정된 구문의 사전 발화와 매칭한다고 결정하는 것을 포함한다. 예를 들어, 음성 지원 장치(125)는 발화가 핫워드 "오케이 컴퓨터"를 포함한다고 결정할 수 있고, 이에 대한 응답으로, 발화 내의 "오케이 컴퓨터"의 화자 식별 특징이 "오케이 컴퓨터"라고 말하는 "아빠"에 대해 이전에 저장된 화자 식별 특징과 매칭하는지 여부를 결정할 수 있다.
- [0055] 일부 구현예들에서, 알려진 사용자들 중 특정 알려진 사용자에 의해 발언된 것으로 발화를 분류하는 것은 화자에 대응하는 시각적 정보(시각 정보)가 특정 알려진 사용자에 대응하는 시각적 정보와 매칭한다고 결정하는 것을 포함한다. 예를 들어, 음성 지원 장치(125)는 카메라를 포함할 수 있고, 카메라를 사용하여 화자의 이미지를 캡처하고, 화자의 지문, 망막 스캔, 얼굴 또는 자세 중 하나 이상이 "아빠"의 해당 시각적 정보와 매칭하는지 여부를 결정한다.
- [0056] 프로세스(200)는 발화의 표현 및 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 단계(220)를 포함한다. 예를 들어, 음성 지원 장치(125)는 "오케이 컴퓨터"라는 오디오 표현 및 "아빠"가 화자라는 표시를 포함하는 쿼리를 서버(155)에 제공할 수 있다.
- [0057] 일부 구현예들에서, 발화의 표현 및 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 단계는 인증 토큰 및 발화가 특정 알려진 사용자가 말한 것으로 분류되었음을 나타내는 플래그를 포함하는 쿼리를 제공하는 단계를 포함한다. 예를 들어, 음성 지원 장치(125)는, 음성 지원 장치(125)에 저장된 모든 인증 토큰 및 "아빠"의 인증 토큰에 대한 "베스트-화자"의 플래그에 대한 "참"의 표시를 포함하는 쿼리를 서버(155)에 제공할 수 있다.
- [0058] 일부 구현예들에서, 발화의 표현 및 화자로서 특정 알려진 사용자의 표시를 포함하는 쿼리를 제공하는 것은 특정 알려진 사용자의 인증 토큰만을 포함하는 쿼리를 제공하는 것을 포함하며, 쿼리에 특정 알려진 사용자의 인증 토큰을 포함하는 것은 발화가 특정 알려진 사용자에 의해 발언된 것으로 분류되었음을 나타낸다. 예를 들어, 음성 지원 장치(125)는 "아빠"에 대한 단일 인증 토큰을 포함하는 쿼리를 서버(155)에 제공할 수 있다.
- [0059] 도 3은 게스트 사용자로부터 발화를 처리하기 위한 프로세스(300)의 예를 예시하는 흐름도이다. 프로세스(300)의 동작은 도 1a-1c의 시스템(100)과 같은 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있다.
- [0060] 프로세스(300)는 장치의 알려진 사용자에 대응하는 인증 토큰을 저장하는 단계(305)를 포함한다. 예를 들어, 음성 지원 장치(125)는 알려진 사용자 "엄마", "아빠", "매트" 및 "케이티"에 대한 인증 토큰을 수신 및 저장할 수 있다.
- [0061] 프로세스(300)는 화자로부터 발화를 수신하는 단계(310)를 포함한다. 예를 들어, 음성 지원 장치(125)는, 음성

지원 장치(125)의 마이크를 통해 알려진 사용자 "아빠"로부터 "오케이 컴퓨터, 내 일정은?"이라는 발화를 수신할 수 있다.

- [0062] 프로세스(300)는 알려진 사용자들에 의해 발언되지 않은 것으로 발화를 분류하는 단계(315)를 포함한다. 예를 들어, 음성 지원 장치(125)는 "엄마", "아빠", "매트" 또는 "케이티"가 말하지 않은 것으로 발화를 분류할 수 있다.
- [0063] 프로세스(300)는 발화의 표현을 포함하고 그리고 화자로서 특정 알려진 사용자를 나타내지 않는 쿼리를 제공하는 단계(320)를 포함한다. 예를 들어, 음성 지원 장치(125)는 "오케이 컴퓨터"의 오디오 표현을 포함하고 그리고 "엄마", "아빠", "매트" 또는 "케이티"의 표시를 화자로서 표시하지 않는 쿼리를 서버(155)에 제공할 수 있다.
- [0064] 일부 구현예에서, 발화의 표현을 포함하고 그리고 화자로서 알려진 사용자들 중 특정 알려진 사용자가 발언한 것으로 분류되었음을 나타내지 않는 쿼리를 제공하는 단계는 인증 토큰의 풀을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자가 말한 것으로 분류됨을 나타내는 플래그를 포함하지 않는 쿼리를 제공하는 단계를 포함한다. 예를 들어, 음성 지원 장치(125)는, 음성 지원 장치(125)에 저장된 모든 인증 토큰 및 알려진 사용자들의 모든 인증 토큰에 대한 "베스트-화자"의 모든 플래그에 대한 "거짓"의 표시를 포함하는 쿼리를 서버(155)에 제공할 수 있다.
- [0065] 일부 구현예에서, 발화의 표현을 포함하고 그리고 발화가 알려진 사용자들 중 특정 알려진 사용자에 의해 발언된 것으로 분류됨을 나타내지 않는 쿼리를 제공하는 것은 알려진 사용자들의 인증 토큰을 포함하지 않는 쿼리를 제공하는 것을 포함한다. 예를 들어, 음성 지원 장치(125)는 "엄마", "아빠", "매트" 또는 "케이티"에 대한 인증 토큰을 포함하지 않는 쿼리를 서버(155)에 제공할 수 있다. 실제로, 일부 실시 예들에 따른 그러한 경우에, 음성 지원 장치(125)는 발화의 표현을 토큰없이 서버(155)에 제공할 수 있다. 다른 실시 예들에 따르면, 사용자 인증 토큰 이외의 토큰은 장치 인증 토큰과 같은 발화의 표현으로 서버(155)에 제공될 수 있다.
- [0066] 도 4는 음성 지원 장치로부터의 발화를 처리하기 위한 프로세스(400)의 예를 나타내는 흐름도이다. 프로세스(400)의 동작은 도 1a-1c의 시스템(100)과 같은 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있다.
- [0067] 프로세스(400)는 쿼리가 특정 알려진 사용자를 화자로서 나타내는지 여부를 결정하는 단계(405)를 포함한다. 예를 들어, 서버(155)는 쿼리가 알려진 사용자 "아빠"가 쿼리에 대한 화자임을 나타내는 것으로 결정할 수 있다. 일부 구현예들에서, 쿼리가 특정 알려진 사용자를 화자로서 나타내는지 여부를 결정하는 단계는 발화가 특정한 알려진 사용자가 발언한 것으로 분류됨을 나타내는 플래그를 포함하는지 여부를 결정하는 단계를 포함한다. 예를 들어, 서버(155)는 알려진 사용자 "아빠"의 인증 토큰에 대한 "베스트-화자"의 플래그가 "참"으로 설정되었는지 여부를 결정할 수 있다. 일부 구현예들에서, 쿼리가 특정 알려진 사용자를 화자로 표시하는지를 결정하는 단계는 쿼리가 알려진 사용자의 인증 토큰을 포함하는지 여부를 결정하는 단계를 포함한다. 예를 들어, 서버(155)는 쿼리가 알려진 사용자 "아빠"에 대한 인증 토큰을 포함한다고 결정할 수 있다.
- [0068] 프로세스(400)는 쿼리가 특정 알려진 사용자를 화자로 표시하는지 여부를 결정하는 것에 응답하여, 쿼리가 알려진 사용자의 것이 아니라고 결정하는 단계(410)를 포함한다. 예를 들어, 서버(155)는 특정 알려진 사용자가 쿼리에 대한 화자가 아니라고 결정하면 쿼리는 게스트 사용자의 쿼리인 것으로 결정할 수 있다.
- [0069] 프로세스(400)는 쿼리가 알려진 사용자의 것이 아니라고 결정하는 것에 응답하여, 쿼리가 비-개인적인지 여부를 결정하는 단계(415)를 포함한다. 예를 들어, 쿼리가 알려진 사용자의 것이 아니라고 결정한 것에 응답하여, 서버(155)는 발화 "오케이 컴퓨터, 내 일정은?"에 대한 쿼리가 개인적이라고 결정할 수 있다. 다른 예에서, 쿼리가 알려진 사용자의 것이 아니라고 결정한 것에 응답하여, 서버(155)는 발화에 대한 쿼리 "오케이 컴퓨터, 몇시입니까?"가 비-개인적인 것으로 결정할 수 있다. 일부 구현예들에서, 쿼리가 비-개인적인지 여부를 결정하는 것은 쿼리에 대한 응답(답변)이 개인 정보에 의존하는지(예를 들어, 보안 사용자 계정에 대한 액세스를 요구함)를 결정하는 것을 포함한다. 예를 들어, 서버(155)는 "내 일정은?"에 대한 쿼리가, 응답(답변)이 사용자가 누구인지와 해당 사용자의 개인 정보에 대한 액세스에 의존하기 때문에, 개인적인 것으로 결정할 수 있다. 다른 예에서, 응답이 사용자가 누구인지에 의존하지 않기 때문에 "몇시입니까?"라는 쿼리가 비-개인적이라고 결정할 수 있다.
- [0070] 프로세스(400)는 쿼리가 비-개인적인 것이라는 결정에 응답하여, 쿼리에 대한 응답을 제공하려고 시도하는 단계(420)를 포함한다. 예를 들어, "몇시입니까?"라는 쿼리가 비-개인적인 것으로 결정한 것에 응답하여, 서버(155)는 현재 시간을 결정하려고 시도하고 음성 지원 장치(125)에 "오전 11시입니다"라는 응답을 제공할 수 있다.

- [0071] 도 5는 음성 지원 장치로부터의 발화를 처리하기 위한 프로세스(500)의 예를 나타내는 흐름도이다. 프로세스(500)의 동작은 도 1a-1c의 시스템(100)과 같은 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있다.
- [0072] 프로세스(500)는 쿼리를 수신하는 단계(510)를 포함한다. 예를 들어, 서버(155)는 (i) 발화 "오케이 컴퓨터, 내 일정은?"의 표현, (ii) 알려진 사용자들에 대한 인증 토큰들의 풀, 및 (iii) 알려진 사용자 "아빠"가 발화를 발 언했다는 표시를 수신할 수 있다.
- [0073] 프로세스(500)는 쿼리가 개인 데이터를 요구하는지 결정하는 단계(512)를 포함한다. 예를 들어, 서버(155)는 "내 일정은?"이 화자의 개인 데이터를 요구한다고 결정할 수 있다.
- [0074] 프로세스(500)가, 쿼리가 개인 데이터를 요구하지 않는다고 결정하면, 프로세스(500)는 쿼리가 알려진 사용자 표시를 포함하는지를 결정하는 단계(514)를 포함한다. 예를 들어, 서버(155)는 "아빠"의 인증 토큰에 대해 "베 스트-화자"의 플래그가 참으로 설정되어 있다고 결정할 수 있다.
- [0075] 프로세스(500)가 쿼리가 알려진 사용자 표시를 포함한다고 결정하면, 그 후 프로세스(500)는 표시된 알려진 사용자에 대한 인증 토큰을 사용하여 로그인 또는 다른 자격증명(credential) 동작을 수행하고(516), 데이터를 검색하고(518), 음성 지원 장치에 제공(serve)한다(520). 예를 들어, 서버(155)는 "아빠"에 대한 인증 토큰을 사용하여 "아빠"의 계정에 로그인하고, "아빠"의 일정을 검색하고, 검색된 일정에 기초하여 "아빠"에 대한 약속의 요약과 함께 응답을 음성 지원 장치(125)에 제공한다.
- [0076] 514 단계로 돌아가서, 프로세스(500)가 대신 쿼리가 알려진 사용자 표시를 포함하지 않는다고 결정하면, 프로세스는 "액세스 안됨" 메시지를 제공하는 단계(522)를 포함한다. 예를 들어, 서버(155)는 음성 지원 장치(125)에 "일정에 액세스할 수 없습니다"라는 응답을 제공할 수 있다. "액세스 안됨" 메시지를 제공한 후, 프로세스(500)는 선택적으로 등록 지원을 제공하는 단계(524)를 포함할 수 있다. 예를 들어, 서버(155)는 "원하실 경우 등록 프로세스를 안내할 수 있습니다"라는 응답을 음성 지원 장치(125)에 제공할 수 있다.
- [0077] 512 단계로 돌아가서, 프로세스(500)가 대신 쿼리가 개인 데이터를 요구하지 않는다고 결정하면, 프로세스는 쿼리가 다른 등록을 요구하는지를 결정한다(526). 예를 들어, 서버(155)는 가입(subscription) 기반 오디오 서비스에 액세스하기 위해 쿼리를 처리하려면 알려진 사용자를 가입에 등록해야 한다고 결정할 수 있다.
- [0078] 프로세스(500)가, 쿼리가 다른 등록을 요구한다고 결정하면, 프로세스(500)는 쿼리가 풀의 하나 이상의 알려진 사용자를 포함하는지를 결정한다(528). 예를 들어, 서버(155)는 풀의 알려진 사용자가 오디오 서비스에 가입하지 않은 것으로 결정할 수 있다. 프로세스(500)가 풀이 하나 이상의 알려진 사용자를 포함하지 않는다고 결정하면, 프로세스는 "액세스 안됨" 메시지를 제공하는 단계(522)를 포함하고 선택적으로 등록을 지원한다(524).
- [0079] 528 단계로 돌아가서, 프로세스(500)가, 쿼리가 풀의 적어도 하나의 알려진 사용자를 포함한다고 결정하면, 프로세스(500)는 쿼리를 실행하고(530), 음성 지원 장치에 제공(serve)(532)하는 것을 포함한다. 예를 들어, 서버(155)는 가입 오디오 서비스에 액세스하고 그리고 가입 오디오 서비스에 대한 액세스를 사용하여 음성 지원 장치(125)에 응답을 제공한다.
- [0080] 526 단계로 돌아가서, 프로세스(500)가, 쿼리가 다른 등록을 요구하지 않는다고 결정하면, 프로세스(500)는 쿼리를 실행하는 단계(530) 및 음성 지원 장치에 서비스하는 단계(532)를 포함한다. 예를 들어, 서버(155)는 "몇 시입니까?"에 대한 쿼리가 다른 등록을 요구하지 않는다고 결정하고, 시간을 나타내는 응답을 제공한다.
- [0081] 쿼리가 개인 데이터를 요구하는지를 결정하는 단계(512)가 프로세스(500)에서 먼저 도시되지만, 프로세스(500)는 상이할 수 있다. 예를 들어, 프로세스(500)는 쿼리가 알려진 사용자 표시를 포함하는지를 결정(514)하는 대신에, 알려진 사용자가 표시되지 않으면 쿼리가 개인 데이터를 요구하는지를 결정한다(512).
- [0082] 도 6은 화자 식별에 사용되는 샘플들의 세트(샘플 세트)를 업데이트하는 시스템(600)의 예를 나타내는 블록도이다. 시스템(600)은 핫워드 검출기(610), 샘플 선택기(620), 레퍼런스 샘플(reference sample) 데이터 저장소(630) 및 최근 샘플(recent sample) 데이터 저장소(640)를 포함한다. 일부 구현예들에서, 핫워드 검출기(610)는 시스템(100)에 대해 설명된 음성 지원 장치(125)와 같은 음성 지원 장치 상에서 구현될 수 있고, 샘플 선택기(620), 레퍼런스 샘플 데이터 저장소(630) 및 최근 샘플 데이터 저장소(640)는 시스템(100)에 대해 설명된 서버(155)와 같은 서버에서 구현된다. 예를 들어, 음성 지원 장치(125)는 핫워드 검출기(610)를 포함할 수 있고, 음성 지원 장치(125)는 특정 알려진 사용자가 핫워드를 말했다는지 여부를 결정하기 위해 핫워드 검출기(610)를 사용할 수 있다. 다른 예에서, 핫워드 검출기(610) 및 샘플 선택기(620)는 음성 지원 장치(125) 상에 위치될 수 있다.

- [0083] 특정 알려진 사용자의 등록 후 핫워드를 말하는 특정 알려진 사용자의 추후 검출을 개선하기 위한 일부 구현예들에서, 시스템(100)은 사용자가 나중에 쿼리를 제공할 때 핫워드를 말하면 추가적인 화자 식별 특징들을 저장할 수 있다. 예를 들어, 알려진 사용자 "아빠"는 초기 등록 동안 "오케이 컴퓨터"라고 세 번 말할 수 있고, 음성 지원 장치(125)는 이들 세 번의 화자 식별 특징을 저장할 수 있다. 알려진 사용자 "아빠"가 "오케이 컴퓨터"라고 말함으로써 쿼리를 제공할 때 등록 후, 음성 지원 장치(125)는 그 쿼리에 대해 "오케이 컴퓨터"라는 말을 사용하는 것에 대해 추가 화자 식별 특징을 저장할 수 있다. 계속해서 예를 들어, 음성 지원 장치(125)는 알려진 사용자 "아빠"가 나중에 "오케이 컴퓨터"라고 말할 때를 결정하기 위해 4세트의 화자 식별 특징 모두를 사용할 수 있다.
- [0084] 화자 식별 특징의 사용이 여기에서 일반적으로 설명되지만, 오디오 기록이 유사하게 사용될 수 있다. 예를 들어, 음성 지원 장치(125)는 "오케이 컴퓨터"라고 말하는 알려진 사용자 "아빠"에 대응하는 4개의 오디오 기록을 저장한 다음, 4개의 오디오 기록을 사용하여 나중에 핫워드를 말하는 알려진 사용자 "아빠"를 검출하는데 사용될 수 있는 핫워드 검출 모델을 생성할 수 있다. 핫워드 검출 모델은 4개의 오디오 기록으로부터 추출된 화자 식별 특징에 기초하여 생성될 수도 있다. 따라서, 핫워드를 말하고 있는 알려진 사용자를 검출하기 위해 화자 식별 특징을 저장하고 사용하는 시스템(100)의 설명은 오디오 기록을 저장하고 사용하여 알려진 사용자가 핫워드를 말하고, 그 반대로도 검출하는 것에 유사하게 적용될 수 있다.
- [0085] 시스템(100)이 추가 화자 식별 특징을 사용하게 함으로써 시스템(100)은 핫워드를 말하는 알려진 사용자의 더 많은 샘플을 수집할 수 있기 때문에 화자가 핫 워드를 말할 때를 보다 정확하게 검출할 수 있다. 추가로 또는 대안적으로, 등록 후 핫워드를 말하는 사용자의 추가 화자 식별 특징을 사용하면 시스템(100)이 사용자가 핫 워드를 말하는 방식의 변화에 적응할 수 있다. 예를 들어, 핫워드를 발음하는 사용자의 음성 또는 방법은 시간이 지남에 따라 천천히 변할 수 있다.
- [0086] 일부 구현예들에서, 시스템(100)은 알려진 사용자가 핫워드를 말하는 제한된 횟수의 화자 식별 특징들을 사용할 수 있다. 예를 들어, 음성 지원 장치(125)는 최대 8개, 10개, 15개 또는 다른 소정 개수의 오디오 기록 및 해당 화자 식별 특징만을 사용할 수 있다. 음성 지원 장치(125)는 등록 동안 그리고 미리 결정된 수의 오디오 레코딩의 나머지 동안, 핫워드를 말하고 있는 알려진 사용자로부터 화자 식별 특징을 항상 사용할 수 있고, 쿼리에 대한 핫워드를 말하는 사용자의 최신 오디오 기록을 사용할 수 있다. 예를 들어, 음성 지원 장치(125)는 검출을 위해 총 8개의 오디오 기록만을 사용할 수 있으므로, 등록 동안 핫워드를 말하는 알려진 사용자의 3개의 오디오 기록을 모두 사용하고, 알려진 사용자가 핫워드를 말한 가장 최근 5개의 오디오 기록을 사용한다.
- [0087] 그러나, 다수의 사용자가 음성 지원 장치(125)를 사용할 수 있기 때문에, 시스템(100)이 알려진 사용자가 핫워드를 말한 것을 높은 신뢰도로 결정할 수 있다면, 시스템(100)은 핫워드를 말하고 있는 알려진 사용자의 추후 검출을 위해 화자 식별 특징만을 사용하기로 결정할 수 있다. 예를 들어, 음성 지원 장치(125)는 "오케이 컴퓨터, 음악 재생"의 특정 발화에 대해 음악을 재생한다"에 대해, "오케이 컴퓨터"는 특정 알려진 사용자가 핫워드를 말할 것으로 예상되는 것과 완전히 동일하지 않다고 결정할 수 있다. 이에 응답하여, 음성 지원 장치(125)는 특정 알려진 사용자가 발화를 말했다 가능성이 얼마나 높은지를 반영하는 신뢰도 점수를 결정할 수 있고, 이는 알려진 사용자가 좋아하는 음악을 재생하기에 충분히 높지만, 음성 지원 장치(125)가 알려진 사용자가 말하고 있는 핫워드를 나중에 검출하기 위해 사용하기에는 충분히 높지 않다.
- [0088] 사용자가 나중에 검출하기 위해 사용하기로 결정한 화자 식별 특징들에 대해, 음성 지원 장치(125)는 화자 식별 특징들 또는 오디오 기록들이 획득되었을 때의 표시로 화자 식별 특징들 또는 오디오 기록들에 라벨을 붙일 수 있다. 예를 들어, 음성 지원 장치(125)는 타임스탬프로 오디오 기록에 라벨을 붙일 수 있고 그리고 타임스탬프를 사용하여 어떤 오디오 기록이 검출에 사용하기 위한 가장 최근의 오디오 기록인지를 결정할 수 있다.
- [0089] 핫워드 검출기(610)는 핫워드가 특정 사용자에게 의해 발언되었을 가능성이 있는지를 결정하고 표시할 수 있다. 예를 들어, 핫워드 검출기(610)는 "오케이 컴퓨터, 내 일정은?"이라고 말하는 알려진 사용자 "아빠"의 오디오 표현을 수신할 수 있고, 알려진 사용자 "아빠"가 핫워드 "오케이 컴퓨터"를 말했다 가능성이 있는 것으로 결정하고, 이에 응답하여, "아빠"의 표시 및 오디오 표현을 샘플 선택기(620)에 제공한다.
- [0090] 음성 지원 장치(125)와 관련하여 위에서 설명된 바와 같이, 핫워드 검출기(610)는 특정 알려진 사용자가 핫워드를 말했다는 여부를 검출하기 위해 샘플 세트, 예를 들어 하나 이상의 화자 식별 특징 또는 오디오 기록(레코딩)을 사용할 수 있다. 예를 들어, 핫워드 검출기(610)는 레퍼런스 샘플 A-C 및 최근 샘플 E-I을 사용하여 "오케이 컴퓨터"를 말하는 알려진 사용자 "아빠"를 인식하도록 훈련된 신경망 모델일 수 있다. 레퍼런스 샘플은 등록 프로세스 동안 알려진 사용자로부터 획득된 샘플 일 수 있고, 최근 샘플은 쿼리로부터 획득된 샘플

일 수 있다.

- [0091] 샘플 선택기(620)는 오디오 표현 및 오디오 표현이 핫워드를 말할 가능성이 있는 알려진 사용자에게 대응한다는 표시를 획득하고, 이에 응답하여, 오디오 표현의 부분을 새로운 최근 샘플로서 저장하고 그 저장된 샘플들로부터 다른 샘플 세트를 획득한다. 예를 들어, 샘플 선택기(620)는 "오케이 컴퓨터, 내 일정은?"의 오디오 표현을 획득하고, "오케이 컴퓨터"에 대응하는 부분으로부터의 샘플을 최근 샘플 J로서 생성한 다음, 새로운 세트에 최근 샘플 E를 포함하지 않고 레퍼런스 샘플 A-C 및 최근 샘플 F-J를 포함하는 새로운 샘플 세트를 선택할 수 있다.
- [0092] 샘플 선택기(620)는 레퍼런스 샘플 데이터 저장소(630)의 레퍼런스 샘플들 및 최근 샘플 데이터 저장소(640)의 최근 샘플들로부터 다른 샘플 세트를 획득할 수 있다. 예를 들어, 샘플 선택기(620)는 레퍼런스 샘플 데이터 저장소(630)로부터 레퍼런스 샘플 A-C를 획득하고, 최근 샘플 데이터 저장소(640)로부터 최근 샘플 E-J를 획득한 다음, 레퍼런스 샘플 A-C 및 최근 샘플 F-J를 선택할 수 있다.
- [0093] 서버(155)에 대해 전술한 바와 유사하게, 샘플 선택기(620)는 샘플이 얼마나 최근인지에 기초하여 샘플 세트에서 사용할 샘플을 선택할 수 있다. 일부 구현예들에서, 알려진 사용자가 말한 핫워드의 검출에 필요한 데이터 전송, 저장 및 프로세싱의 양을 감소시키기 위해, 핫워드 검출기(610)는 검출을 위해 제한된 수의 샘플만을 사용할 수 있다. 예를 들어, 핫워드 검출기(610)는 검출을 위해 8개의 샘플만을 사용할 수 있다. 그러나, 알려진 사용자의 핫워드를 말하는 방식의 변경, 예를 들어, 사용자가 아프거나, 사용자가 시끄러운 환경에 있고, 사용자의 억양이 변경된 경우, 샘플 선택기(620)는 가장 최근에 획득된 샘플을 선택할 수 있다. 또한, 레퍼런스 샘플은 알려진 사용자가 쿼리를 제공하기 위해 핫워드를 말하고 있는 가장 정확한 표현을 갖는 것으로 간주될 수 있지만, 샘플 선택기는 미리 결정된 개수의 레퍼런스 샘플, 예를 들어, 3개 또는 모두를 선택할 수 있고, 및 가장 최근에 획득된 최근 샘플들 중 미리 결정된 개수를 선택할 수 있다.
- [0094] 일부 구현예들에서, 샘플 선택기(620)는 샘플이 생성된 시간, 예를 들어 타임스탬프의 표시로 생성될 때 각각의 샘플을 라벨링함으로써 가장 최근에 획득된 최근 샘플을 결정할 수 있고, 최근 샘플들을 상기 표시와 함께 최근 샘플 데이터 저장소(640)에 저장한다.
- [0095] 도 6에 도시된 바와 같이, 샘플 선택기(620)는 "오케이 컴퓨터, 내 일정은?"으로부터 최근 샘플 J를 생성한 후, 샘플 선택기(620)는 레퍼런스 샘플 데이터 저장소(630)로부터 레퍼런스 샘플 A-C를 획득하고, 최근 샘플 데이터 저장소(640)로부터 최근 샘플 E-I를 획득하고, 레퍼런스 샘플들 모두를 선택하고, 최근 샘플들 E-J 중에서 가장 최근에 생성된 5개의 최근 샘플들에 대한 결정에 기초하여 최근 샘플들 F-J를 선택함으로써, 레퍼런스 샘플 A-C 및 최근 샘플 F-J를 포함하는 세트를 획득할 수 있다.
- [0096] 일부 구현예들에서, 핫워드 검출기(610)는 알려진 사용자가 핫워드를 말한 것에 대한 신뢰도의 표시를 부가적으로 또는 대안적으로 제공할 수 있고 그리고 샘플 선택기는 검출에 사용된 샘플 세트에 대해 최신 샘플을 사용하지 여부를 결정할 때 그 신뢰도를 고려할 수 있다. 예를 들어, 샘플 선택기(620)는 95%의 신뢰도 임계 값을 사용할 수 있으므로, 핫워드 검출기(610)가 알려진 사용자가 오디오 표현에서 핫워드를 말한 것으로 95% 미만의 신뢰도를 나타내는 경우 오디오 표현으로부터 새로운 샘플을 생성하지 않을 수 있다.
- [0097] 샘플 선택기(620)는 오디오 표현이 어떤 유형의 장치로부터 왔는지 추가적으로 또는 대안적으로 결정할 수 있고, 그 유형의 표시로 최근 샘플을 라벨링한 다음, 새로운 샘플 세트를 생성할 때 상기 유형을 고려할 수 있다. 예를 들어, 샘플 선택기(620)는 테이블-탑(table-top) 음성 지원 장치가 가장 최신의 샘플 세트를 요청하고 있음을 결정하고, 이에 응답하여, 테이블-탑 음성 지원 장치로부터 나온 것으로 라벨링된 최근 샘플들 중에서만 최근 샘플들을 선택할 수 있다. 다른 예에서, 샘플 선택기(620)는 텔레비전 음성 지원 장치가 최신 샘플 세트를 요청하고 있다고 결정할 수 있고, 이에 응답하여, 텔레비전 음성 지원 장치로부터 가장 최근의 2개의 샘플과 아직 선택되지 않은 최근의 샘플들 중 3개의 가장 최신 샘플을 결정하고 선택할 수 있다.
- [0098] 추가적으로 또는 대안적으로, 핫워드 검출기는 음성 지원 장치에서 구현되고 샘플 선택기(620)는 서버에서 구현되며, 서버는 샘플 세트에 기초하여 검출을 위한 모델을 생성한 다음 샘플 자체를 음성 지원 장치에 제공하지 않고 상기 모델을 음성 지원 장치에 전송할 수 있다. 예를 들어, 서버(155)는 레퍼런스 샘플 A-C 및 최근 샘플 F-J를 포함하는 샘플 세트를 획득하고, 상기 샘플 세트를 사용하여 핫워드 검출 모델을 학습시킨 후, 학습된 핫워드 검출 모델을 음성 지원 장치(125)에 전송할 수 있다.
- [0099] 일부 구현예들에서, 시스템(100)은, 시스템(100)이 핫워드를 말하고 있는 특정 알려진 사용자의 화자 식별 특징을 이미 저장 한 경우, 음성 가능 장치(125)가 핫워드를 말하지 않고 핫워드를 말하고 있는 특정 알려진 사용자

의 화자 식별 특징을 획득할 수 있게 할 수 있다. 예를 들어, 사용자가 음성 지원 장치(125)와 함께 사용할 계정에 대한 로그인 정보를 제공한 후 초기 등록 동안, 음성 지원 장치(125)는 서버(155)가 다른 음성 지원 장치에 사전 등록하는 동안 "오케이 컴퓨터"라고 3번 말한 사용자의 오디오 기록이 저장되어 있다고 결정하고, 이에 대한 응답으로, 등록시에 특정 알려진 사용자에게 "오케이 컴퓨터"라고 3번 말하도록 하는 대신 서버(155)로부터 사전 등록에 대한 오디오 기록을 요청할 수 있다.

[0100] 저장된 화자 식별 특징을 사용함으로써, 시스템(100)은 사용자가 추가 음성 지원 장치에 등록하기 위해 핫워드를 여러 번 말할 필요없이 추가 음성 지원 장치를 사용하여 사용자가 신속하게 시작할 수 있게 할 수 있다. 후속 장치들에 대해, 시스템(100)은 제1 음성 지원 장치로의 등록으로부터 저장된 "오케이 컴퓨터"라고 말하는 사용자의 이전 오디오 기록들로부터 화자 식별 특징들을 대신 사용할 수 있다. 또한, 다양한 음성 지원 장치가 모두 동일한 화자 식별 특징을 사용할 수 있기 때문에, 시스템(100)은 사용자가 핫워드를 말할 때 다양한 장치가 일관성 있고 균일하게 검출되도록 할 수 있다.

[0101] 특정 예에서, 사용자는 스마트폰 상의 컴패니언 애플리케이션을 통해 음성 지원 장치(125)를 사용하도록 등록할 수 있고, 사용자가 음성 지원 장치(125)의 알려진 사용자로서 등록하고 싶다는 것을 표시하고, 컴패니언 애플리케이션에서 사용자 계정에 대한 로그인 자격증명(credentials)을 제공하고, 컴패니언 애플리케이션은 서버(155)가 사용자의 계정에 핫워드 "오케이 컴퓨터"를 말하는 사용자의 오디오 기록이 저장되어 있는지 여부를 결정하도록 요청할 수 있다. 서버(155)는 저장된 오디오 기록, 예를 들어 부울 "참"이 존재한다는 표시로 응답할 수 있고, 컴패니언 애플리케이션은, 이에 응답하여, 사용자에게 핫워드를 여러 번 말하도록 요청하는 스크린을 스킵하고, 대신에 서버(155)에 오디오 기록을 요청하는 명령을 음성 지원 장치(125)에 전송할 수 있다. 이에 응답하여, 서버(155)는 오디오 기록을 음성 지원 장치(125)로 전송할 수 있다.

[0102] 일부 구현예들에서, 음성 지원 장치(125)가 알려진 사용자에 의해 발언되는 핫워드의 추후 검출을 위한 쿼리의 일부로서 핫워드를 말하는 알려진 사용자로부터의 추가적인 화자 식별 특징을 어떻게 사용할 수 있는지와 유사하게, 시스템(100)의 음성 지원 장치는 이러한 추가적인 화자 식별 특징을 서로 공유하여 모든 음성 지원 장치가 핫워드를 말하는 알려진 사용자의 검출을 향상시킬 수 있다. 예를 들어, 사용자 "아빠"는 음성 지원 장치(125)에 "오케이 컴퓨터, 오늘 나의 약속은 무엇인가"라고 말할 수 있고, 음성 지원 장치(125)는 나중에 "오케이 컴퓨터"라고 말하는 사용자 "아빠"를 검출하도록 사용하기 위해 제2 음성 지원 장치에 제공되는 오디오 기록을 생성할 수 있다.

[0103] 일부 구현예들에서, 음성 지원 장치들은 서버(155)를 통해 추가적인 화자 식별 특징들을 공유할 수 있다. 예를 들어, 음성 지원 장치(125)가 핫워드 발화에 대한 신뢰도 점수가 미리 정해진 신뢰도 점수 임계 값을 만족한다고 결정할 때마다, 음성 지원 장치(125)는, 음성 지원 장치(125)에 저장된 특정 알려진 사용자에 대한 토큰과 함께 핫워드를 말하는 알려진 사용자의 화자 식별 특징 또는 오디오 기록을 서버(155)에 제공할 수 있다. 서버(155)는 해당 알려진 사용자 계정을 식별하고, 토큰의 포함에 기초하여 오디오 기록이 알려진 사용자 계정에 대해 저장되어야 한다는 것을 입증할 수 있고, 그 다음 화자 식별 특징 또는 오디오 기록을 서버(155)에 저장하고, 알려진 사용자가 사용하는 것으로 등록한 다른 모든 음성 지원 장치에 화자 식별 특징 또는 오디오 기록을 제공한다. 다른 음성 지원 장치는 핫워드를 말하고 있는 알려진 사용자의 추후 검출을 위해 화자 식별 특징 또는 오디오 기록을 추후 사용할 수 있다.

[0104] 일부 구현예들에서, 서버(155)는 음성 지원 장치들로부터의 요청들에 응답하여 또는 음성 지원 장치들로부터 요청들을 수신하지 않고 추가적인 화자 식별 특징들 또는 오디오 기록들을 제공할 수 있다. 예를 들어, 음성 지원 장치(125)의 전원이 켜질 때마다(또는 소정의 간격, 예를 들어 매시간, 매일, 매주 등 마다), 음성 지원 장치(125)는 검출에 사용하기 위해 오디오 기록에 대한 임의의 업데이트를 요청하는 요청을 서버(155)에 제공할 수 있다. 다른 예에서, 서버(155)는, 서버(155)가 핫워드를 말하는 특정 알려진 사용자의 새로운 오디오 기록을 저장할 때마다, 특정 알려진 사용자의 표시로 라벨링된 추가 오디오 기록을 푸시할 수 있다.

[0105] 일부 구현예들에서, 시스템(100)은 전체 세트의 화자 식별 특징들 또는 오디오 기록들을 음성 지원 장치들로 전송할 수 있다. 예를 들어, 서버(155)는 "오케이 컴퓨터"라고 말하는 알려진 사용자를 검출하기 위해 오디오 기록에 대한 요청을 수신할 수 있고, 이에 대한 응답으로, 알려진 사용자의 등록으로부터 3개의 오디오 기록 및 "오케이 컴퓨터"라고 말하는 알려진 사용자의 가장 최근의 5개의 오디오 기록을 쿼리의 일부로서 제공할 수 있다.

[0106] 일부 구현예들에서, 시스템(100)은 음성 지원 장치들이 이미 저장한 화자 식별 특징들 또는 오디오 기록들의 전송을 회피함으로써 대역폭 사용량을 감소시키려고 시도할 수 있다. 예를 들어, 서버(155)는 특정 음성 지원 장

치료부터 현재의 오디오 기록 세트에 대한 요청을 수신하고 그리고 8개의 오디오 기록 각각에 대한 식별자, 예를 들어 고유한 영숫자 텍스트를 전송할 수 있다. 특정 음성 지원 장치는 그 식별자를 음성 지원 장치가 이미 저장한 오디오 기록의 식별자와 비교하고 그리고 음성 지원 장치가 저장하지 않은 식별자에 대해서만 오디오 기록을 요청할 수 있다. 서버(155)는 음성 지원 장치가 식별자에 의해 요청한 오디오 기록을 제공할 수 있고, 그 후, 음성 지원 장치는 핫워드를 말하는 알려진 사용자의 추후 검출을 위해 그 오디오 기록의 세트를 사용할 수 있다.

[0107] 부가적으로 또는 대안적으로, 시스템(100)은 음성 지원 장치가 특정 식별 특징 또는 오디오 기록을 요구할 필요가 없도록 각 음성 지원 장치가 어떤 오디오 기록을 저장했는지 추적할 수 있다. 예를 들어, 서버(155)는 특정 음성 지원 장치로부터 현재의 오디오 기록의 세트에 대한 요청을 수신할 수 있고, 특정 음성 지원 장치가 사용할 8개의 오디오 기록들의 세트에 대해 2개의 새로운 오디오 기록을 모두 포함한다고 결정하고, 이에 응답하여, 음성 지원 장치에 2개의 새로운 오디오 기록과 함께 8개의 오디오 기록에 대한 식별자들의 세트를 제공한다. 그 후, 음성 지원 장치는 식별자들의 세트 내의 임의의 식별자들과 매칭하지 않는 식별자들을 갖는 임의의 저장된 오디오 기록을 삭제하고, 핫워드를 나중에 검출하기 위해 상기 세트에 대응하는 식별자들을 갖는 오디오 기록을 사용할 수 있다.

[0108] 일부 구현예들에서, 발화에 대한 신뢰도 점수는 전적으로 화자 식별 특징들 또는 오디오 기록들에 기초할 수 있다. 예를 들어, 음성 지원 장치(125)는 현재 발화에 대한 화자 식별 특징들을 획득하고, 화자 식별 특징들을 저장된 화자 식별 특징들과 비교하고, 그 비교로부터 직접 신뢰도 점수를 생성할 수 있다.

[0109] 대안적으로, 발화에 대한 신뢰도 점수는 다른 요인들에 기초할 수 있다. 예를 들어, 음성 지원 장치(125)는 다른 쿼리가 방금 수신되었는지를 결정하고, 현재 발화에 대한 초기 신뢰도 점수 및 이전 신뢰도 점수의 평균을 취할 수 있다. 다른 예에서, 음성 지원 장치(125)는 사용자가 최근에 알려진 사용자에 대한 패스코드(passcode)를 제공한 것으로 결정하고, 이에 응답하여 화자 식별 특징 또는 오디오 기록만으로 생성되는 것보다 상기 발화에 대해 더 높은 신뢰도 점수를 생성할 수 있다.

[0110] 일부 구현예들에서, 시스템(100)은 특정 음성 지원 장치의 특성에 기초하여 사용할 특정 음성 지원 장치에 대한 특정 오디오 기록 세트를 결정할 수 있다. 예를 들어, 서버(155)는, 대중에게 공개될 가능성이 있고 그리고 핫워드를 말하는 많은 다른 사람들로부터의 발화에 노출될 수 있는 모바일 컴퓨팅 장치인 음성 지원 장치에 대해 매우 높은 신뢰도를 갖는 오디오 기록만을 사용할 수 있다. 다른 예에서, 서버(155)는 테이블-탑 음성 지원 장치들에 대해 하나의 테이블-탑 음성 지원 장치에 의해 생성된 오디오 기록만을 사용할 수 있다. 이 예에서, 서버(155)는, 서버(155)가 나중에 장치 유형에 기초하여 음성 지원 장치에 사용할 오디오 기록을 결정할 수 있도록, 어떤 유형의 장치가 오디오 기록을 생성했는지를 나타내는 정보로 오디오 기록을 라벨링할 수 있다.

[0111] 도 6의 상기 설명이 알려진 사용자 "아빠"에 대한 샘플을 설명하지만, 핫워드 검출기(610) 및 샘플 선택기(620)는 다수의 다른 알려진 사용자에 대한 샘플을 유사하게 사용할 수 있다. 핫워드 검출기(610)는 상이한 샘플들의 세트를 획득할 수 있고(여기서 각각의 세트는 특정 알려진 사용자에 대한 샘플을 포함함), 알려진 사용자가 그 세트를 사용하여 발화에서 핫워드를 말했는지 여부를 결정할 수 있다. 예를 들어, 핫워드 검출기(610)는 발화를 수신할 수 있고, 동시에 알려진 사용자 "아빠"에 대한 샘플 세트를 사용하여 학습된(훈련된) 제1 핫워드 검출기 모델을 사용할 수 있고, 알려진 사용자 "엄마"에 대한 샘플 세트를 사용하여 학습된 제2 핫워드 검출기 모델을 사용할 수 있고, 제2 핫워드 검출기 모델의 출력이 알려진 사용자 "엄마"가 핫워드 "오케이 컴퓨터"를 말한 것에 대해 97% 신뢰도를 나타내고, 제2 핫워드 검출기 모델의 출력이 알려진 사용자 "아빠"가 핫워드를 말한 것에 대해 8% 신뢰도를 나타내는 것으로 결정할 수 있고, 97% 신뢰도가 미리 결정된 임계 값 95%보다 크고 그리고 8% 신뢰도보다 크다는 결정에 응답하여, 발화가 핫워드를 말하는 알려진 사용자 "엄마"를 포함한다고 결정할 수 있다.

[0112] 도 7은 화자 식별에 사용되는 샘플 세트를 업데이트하기 위한 프로세스(700)의 예를 나타내는 흐름도이다. 프로세스(200)의 동작은 도 6의 샘플 선택기(620) 및 핫워드 검출기(610)와 같은 하나 이상의 컴퓨팅 시스템에 의해 수행될 수 있다.

[0113] 프로세스(700)는 발화가 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플에 기초하여 핫워드를 말하고 있는 특정 사용자를 포함한다고 결정하는 단계(705)를 포함한다. 예를 들어, 핫워드 검출기(610)는 발화가 핫워드 "오케이 컴퓨터"를 말하는 알려진 사용자 "아빠"를 포함하는 것으로 검출하기 위해 최근 샘플 E-I 및 레퍼런스 샘플 A-C의 세트를 사용할 수 있다.

- [0114] 일부 구현예들에서, 발화가 핫워드를 말하고 있는 특정 사용자의 제1 세트의 샘플들에 기초하여 핫워드를 말하는 특정 사용자를 포함한다고 결정하는 단계는 상기 제1 세트의 샘플을 사용하여 핫워드 검출 모델을 생성하고, 핫워드 검출 모델에 발화를 입력하는 단계; 및 핫워드 검출 모델이 발화를 핫워드를 말하는 특정 사용자를 포함하는 것으로 분류했다고 결정하는 단계를 포함할 수 있다. 예를 들어, 핫워드 검출기(610)는 제1 세트의 샘플을 사용하여 핫워드 검출 모델을 생성하고, 핫워드 검출 모델로 발화를 입력한 다음, 핫워드 검출기의 출력으로서 핫워드 검출 모델로부터의 출력을 사용할 수 있다.
- [0115] 프로세스(700)는 발화가 핫워드를 말하는 특정 사용자의 제1 세트의 샘플에 기초하여 핫워드를 말하는 특정 사용자를 포함한다는 결정에 응답하여, 발화의 적어도 부분(일부)을 새로운 샘플로서 저장하는 단계(710)를 포함한다. 예를 들어, 알려진 사용자 "아빠"가 발화에서 핫워드를 발언했다는 표시를 핫워드 검출기(610)로부터 수신한 것에 응답하여, 샘플 선택기(620)는 "오케이 컴퓨터"에 대응하는 발화의 부분을 알려진 사용자 "아빠"에 대한 새로운 최신 샘플 J로서 저장할 수 있다.
- [0116] 프로세스(700)는 발화를 말하는 특정 사용자의 제2 세트의 샘플을 획득하는 단계(715)를 포함하고, 여기서 제2 세트의 샘플은 새로운 샘플을 포함하고 제1 세트의 샘플 내의 모든 샘플보다 적은 샘플을 포함한다. 예를 들어, 샘플 선택기(620)는 최근 샘플 E를 포함하지 않고, 레퍼런스 샘플 A-C 및 최근 샘플 F-J를 포함하는 제2 세트의 샘플을 획득할 수 있다.
- [0117] 일부 구현예들에서, 발화를 말하는 특정 사용자의 제2 세트의 샘플을 획득하는 단계(715)(여기서 제2 세트의 샘플은 새로운 샘플을 포함하고 제1 세트의 샘플 내의 모든 샘플보다 적은 샘플을 포함함)는 미리 결정된 수의 최근 저장된 샘플을 제2 세트의 샘플로서 선택하는 단계를 포함한다. 예를 들어, 샘플 선택기(620)는 가장 최근에 저장된 최근 샘플 중 3개, 5개, 8개 또는 다른 미리 결정된 수를 선택할 수 있다. 일부 구현예에서, 발화를 말하는 특정 사용자의 제2 세트의 샘플을 획득하는 단계(여기서 제2 세트의 샘플은 새로운 샘플을 포함하고 제1 세트의 샘플 내의 모든 샘플보다 적은 샘플을 포함함)는 미리 결정된 개수의 가장 최근에 저장된 샘플들과 레퍼런스 샘플들의 세트를 모두 선택하여 제2 세트의 샘플로서 함께 결합하는 단계를 포함한다. 예를 들어, 샘플 선택기(620)는 가장 최근에 저장된 최근 샘플의 3개, 5개, 8개 또는 다른 수를 선택하고, 하나, 2개, 3개, 모두 또는 일부 다른 수의 레퍼런스 샘플을 선택할 수 있다.
- [0118] 프로세스(700)는 적어도 핫워드를 말하는 사용자의 제2 세트의 샘플에 기초하여 제2 발화가 상기 핫워드를 말하는 특정 사용자를 포함한다고 결정하는 단계(720)를 포함한다. 예를 들어, 핫워드 검출기(610)는 발화가 핫워드 "오케이 컴퓨터"를 말하는 알려진 사용자 "아빠"를 포함한다는 것을 검출하기 위해, 최근 샘플 E 없이, 최근 샘플 F-J 및 레퍼런스 샘플 A-C의 세트를 사용할 수 있다.
- [0119] 일부 구현예들에서, 프로세스(700)는 제2 세트의 샘플들을 획득하는 것에 응답하여, 제2 세트의 샘플에서가 아니라, 제1 세트의 샘플에서 샘플을 삭제하는 단계를 포함한다. 예를 들어, 샘플 선택기(620)는 최근 샘플 E 없이 제2 세트를 선택한 후, 샘플 선택기(620)는 핫워드 검출을 위한 현재 세트가 최근 샘플 E를 포함하지 않는 것으로 결정하고, 이에 응답하여 최근 샘플 데이터 저장소(640)에 최근 샘플 E를 삭제하도록 지시할 수 있다.
- [0120] 일부 구현예들에서, 프로세스(700)는 서버로부터 제2 새로운 샘플을 수신하는 단계 및 서버로부터의 제2 새로운 샘플을 포함하고 그리고 제2 세트의 샘플 내의 모든 샘플보다 적은 제3 세트의 샘플에 기초하여 핫워드를 말하는 특정 사용자를 제3 발화가 포함하는 것으로 결정하는 단계를 포함한다. 예를 들어, 핫워드 검출기(610)는 최근 샘플 K를 수신한 다음, 다른 발화가 레퍼런스 샘플 A-C 및 최근 샘플 G-K에 기초하여 "오케이 컴퓨터"라고 말하는 알려진 사용자 "아빠"를 포함한다고 결정할 수 있다.
- [0121] 일부 구현예들에서, 프로세스(700)는 서버로부터, 제3 세트의 샘플에서 샘플의 표시를 수신하는 단계; 로컬로 저장되지 않은 제3 세트의 샘플에 있는 샘플을 결정하는 단계; 로컬로 저장되지 않은 제3 세트의 샘플 내의 샘플에 대한 요청을 서버에 제공하는 단계; 요청에 응답하여 서버로부터 로컬로 저장되지 않은 샘플을 수신하는 단계를 포함한다. 예를 들어, 핫워드 검출기(610) 또는 음성 지원 장치(125)는 서버(155)에 현재 세트의 샘플(샘플들의 현재 세트)을 요청하고, 이에 응답하여 핫워드 검출기(610) 또는 음성 지원 장치(125)에 대한 현재 세트의 샘플이 레퍼런스 샘플 A-C 및 최근 샘플 G-K를 포함한다는 표시를 수신하고, 레퍼런스 샘플 A-C 및 최근 샘플 G-J가 이미 로컬에 저장되어 있지만 최근 샘플 K가 로컬에 저장되어 있지 않다고 결정하고, 최근 샘플 K에 대한 요청을 서버(155)에 제공하고, 그 요청에 대한 응답으로 최근 샘플 K를 수신할 수 있다.
- [0122] 일부 구현예에서, 프로세스(700)는 상기 음성 지원 장치가 상기 특정 사용자가 핫워드를 말하는지 여부를 검출할 수 있도록 상기 음성 지원 장치에 제1 세트의 샘플을 제공하는 단계를 포함하며, 핫워드를 말하는 특정 사용

자의 제1 세트의 샘플에 기초하여 핫워드를 말하는 특정 사용자를 발화가 포함하는 것으로 결정하는 단계는 상기 특정 사용자가 핫워드를 말했음을 상기 음성 지원 장치가 검출했다는 표시를 수신하는 단계를 포함한다. 예를 들어, 서버(155)는 발화가 음성 지원 장치(125)로부터 오디오 표현을 수신하는 것에 기초하여 핫워드 "오케이 컴퓨터"를 말하는 알려진 사용자 "아빠" 및 화자로서 플래그된 알려진 사용자 "아빠"에 대한 토큰을 포함하는 것으로 결정할 수 있다.

[0123] 일부 구현예들에서, 프로세스(700)는 제1 세트의 샘플들을 사용하여 핫워드 검출 모델을 생성하는 단계 및 핫워드 검출 모델을 음성 지원 장치에 제공하여 음성 지원 장치가 특정 사용자가 핫워드를 말했는지 여부를 검출할 수 있게 하는 단계를 포함하며, 여기서 핫워드를 말하는 특정 사용자의 제1 세트의 샘플에 기초하여 발화가 핫워드를 말하는 특정 사용자를 포함한다고 결정하는 단계는 상기 음성 지원 장치가 상기 특정 사용자가 핫워드를 말했음을 검출했다는 표시를 수신하는 단계를 포함한다. 예를 들어, 서버(155)는 샘플 세트에 기초하여 핫워드 검출 모델을 생성한 다음 핫워드 검출 모델을 음성 지원 장치(125)에 제공할 수 있다.

[0124] 일부 구현예에서, 프로세스(700)는 특정 사용자가 핫워드를 말했는지 여부를 검출하기 위해 현재 세트의 샘플(현재 샘플 세트)에 대한 요청을 음성 지원 장치로부터 수신하는 단계; 음성 지원 장치에 의해 로컬적으로 저장되지 않은 현재 세트의 샘플 내의 샘플을 결정하는 단계; 상기 음성 지원 장치에, 상기 음성 지원 장치에 의해 로컬적으로 저장되지 않은 현재 세트의 샘플(현재 샘플 세트) 내의 샘플 및 상기 현재 세트의 샘플(현재 샘플 세트) 내의 샘플의 표시를 제공하는 단계를 포함한다. 예를 들어, 서버(155)는 음성 지원 장치(125)로부터 현재 샘플 세트에 대한 요청을 수신할 수 있고, 서버(155)가 이전에 음성 지원 장치(125)에 제공한 샘플을 결정 한 후, 서버(155)가 아직 음성 지원 장치(125)에 제공하지 않았다고 서버(155)가 결정한 샘플과 함께 현재 샘플 세트 내의 샘플들에 대한 표시(indication)를 음성 지원 장치(125)에 제공한다.

[0125] 일부 구현예에서, 시스템(600)은 사용자가 새로운 레퍼런스 샘플을 제공하도록 요청할 수 있게 할 수 있다. 예를 들어, 시스템(600)은 사용자로부터 저장된 샘플을 리셋(reset)하라는 요청을 사용자로부터 수신할 수 있다. 시스템(600)은 사용자가 사용하는 스마트폰에서 컴패니언 애플리케이션을 통해 샘플을 리셋하라는 요청을 수신할 수 있고, 상기 요청에 응답하여, 시스템(600)은 사용자를 위해 저장된 모든 레퍼런스 샘플을 삭제하고 그리고 사용자가 핫워드를 여러 번 말하도록 요청하고, 발화로부터 새로운 레퍼런스 샘플을 생성할 수 있다. 시스템(600)은 모든 새로운 레퍼런스 샘플 및 미리 결정된 수의 가장 최근에 저장된 최근 샘플을 포함하는 제3 세트의 샘플을 생성할 수 있다. 다른 예에서, 시스템(600)은 사용자에 의해 사용된 스마트폰 상의 컴패니언 애플리케이션을 통해 샘플을 리셋하라는 요청을 수신할 수 있고; 상기 요청에 응답하여, 시스템(600)은 사용자를 위해 저장된 레퍼런스 및 최근 샘플 모두를 삭제하고, 사용자가 핫워드를 여러 번 말하고 발화로부터 새로운 레퍼런스 샘플을 생성하도록 요청할 수 있다. 이후 시스템(600)은 새로운 레퍼런스 샘플만을 포함하는 제3 세트의 샘플을 생성할 수 있다.

[0126] 도 8은 본 명세서에 기술된 기법들을 구현하는데 사용될 수 있는 컴퓨팅 장치(800)와 모바일 컴퓨팅 장치(850)의 예시를 도시한다. 컴퓨팅 장치(800)는 랩톱, 데스크톱, 워크스테이션, 개인 휴대 정보 단말기, 서버, 블레이드 서버, 메인 프레임 및 다른 적절한 컴퓨터와 같은 다양한 형태의 디지털 컴퓨터들을 나타내기 위한 것이다. 모바일 컴퓨팅 장치(850)는 개인 휴대 정보 단말기, 셀룰러 전화기, 스마트폰 및 다른 유사한 컴퓨팅 장치들과 같은 다양한 형태의 모바일 장치들을 나타내기 위한 것이다. 여기에 도시된 컴포넌트들, 그들의 연결 및 관계, 및 그들의 기능은 단지 예시적인 것을 의미하며, 제한하는 것을 의미하지는 않는다.

[0127] 컴퓨팅 장치(800)는 프로세서(802), 메모리(804), 저장 장치(806), 메모리(804) 및 다수의 고속 확장 포트(810)에 연결되는 고속 인터페이스(808) 및 저속 확장 포트(814) 및 저장 장치(806)에 연결되는 저속 인터페이스(812)를 포함한다. 프로세서(802), 메모리(804), 저장 장치(806), 고속 인터페이스(808) 및 저속 인터페이스(812) 각각은 다양한 버스들을 사용하여 상호연결되며, 공통 마더보드에 또는 적절한 다른 방식으로 장착된다. 프로세서(802)는 메모리(804) 또는 저장 장치(806)에 저장된 명령어들을 포함하는, 컴퓨팅 장치(800) 내에서 실행하기 위한 명령어들을 프로세싱하여, 고속 인터페이스(808)에 연결된 디스플레이(816)와 같은 외부 입/출력 장치에 GUI(graphical user interface)에 대한 그래픽 정보를 디스플레이할 수 있다. 다른 구현예에서, 다수의(복수의) 프로세서들 및/또는 다수의 버스들이 다수의 메모리들 및 다수의 유형의 메모리와 함께, 적절하게 사용될 수 있다. 또한, 다수의 컴퓨팅 장치들은 필요한 동작의 부분들을 제공하는 각 장치와 연결될 수 있다(예를 들어, 서버 뱅크, 블레이드 서버 그룹 또는 멀티 프로세서 시스템).

[0128] 메모리(804)는 컴퓨팅 장치(800) 내에 정보를 저장한다. 일부 구현예에서, 메모리(804)는 휘발성 메모리 유닛 또는 유닛들이다. 일부 구현예에서, 메모리(804)는 비휘발성 메모리 유닛 또는 유닛들이다. 또한, 메모리(804)

는 자기 또는 광학 디스크와 같은 컴퓨터 판독가능 매체의 다른 형태 일 수 있다.

- [0129] 저장 장치(806)는 컴퓨팅 장치(800)에 대한 대형 저장소를 제공할 수 있다. 일부 구현예에서, 저장 장치(806)는 플로피 디스크 장치, 하드 디스크 장치, 광 디스크 장치 또는 테이프 장치, 플래시 메모리 또는 다른 유사한 고체 상태 메모리 장치, 또는 저장 영역 네트워크 또는 다른 구성의 장치를 포함하는 장치의 어레이와 같은 컴퓨터 판독가능 매체이거나 컴퓨터 판독가능 매체를 포함할 수 있다. 명령어들은 정보 캐리어에 저장될 수 있다. 명령어들은 하나 이상의 프로세싱 장치들(예를 들어, 프로세서(802))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법들을 수행한다. 명령어들은 또한 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(804), 저장 장치(806) 또는 프로세서(802)상의 메모리)와 같은 하나 이상의 저장 장치들에 의해 저장될 수 있다.
- [0130] 고속 인터페이스(808)는 컴퓨팅 장치(800)에 대한 대역폭 집중 동작들을 관리하는 반면, 저속 인터페이스(812)는 낮은 대역폭 집중 동작들을 관리한다. 이러한 기능들의 할당은 단지 예시적인 것이다. 일부 구현예에서, 고속 인터페이스(808)는 메모리(804), 디스플레이(816)(예를 들어, 그래픽 프로세서 또는 가속기를 통해) 및 다양한 확장 카드(도시되지 않음)를 수용할 수 있는 고속 확장 포트(810)에 연결된다. 구현예에서, 저속 인터페이스(812)는 저장 장치(806) 및 저속 확장 포트(814)에 연결된다. 다양한 통신 포트(예를 들어, USB, 블루투스, 이더넷, 무선 이더넷)를 포함할 수 있는 저속 확장 포트(814)는 키보드, 포인팅 장치, 스캐너와 같은 하나 이상의 입력/출력 장치 또는 예를 들어 네트워크 어댑터를 통해 스위치 또는 라우터와 같은 네트워킹 장치에 결합될 수 있다.
- [0131] 컴퓨팅 장치(800)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들어, 그것은 표준 서버(820)로서 또는 그러한 서버들의 그룹에서 다수로 구현될 수 있다. 또한, 랩톱 컴퓨터(822)와 같은 퍼스널 컴퓨터에서 구현될 수 있다. 또한, 랙 서버 시스템(824)의 일부로서 구현될 수 있다. 대안적으로, 컴퓨팅 장치(800)로부터의 컴포넌트들은 모바일 컴퓨팅 장치(850)와 같은 모바일 장치(도시되지 않음) 내의 다른 컴포넌트들과 결합될 수 있다. 상기 장치들 각각은 컴퓨팅 장치(800), 모바일 컴퓨팅 장치(850) 중 하나 이상을 포함할 수 있고, 전체 시스템은 서로 통신하는 다수의 컴퓨팅 장치들로 구성될 수 있다.
- [0132] 모바일 컴퓨팅 장치(850)는 다른 여러 컴포넌트들 중에서도 특히, 프로세서(852), 메모리(864), 디스플레이(854)와 같은 입/출력 장치, 통신 인터페이스(866) 및 송수신기(868)를 포함한다. 모바일 장치(850)에는 또한 추가적 저장을 제공하기 위해 마이크로 드라이브 또는 다른 장치와 같은 저장 장치가 제공될 수 있다. 프로세서(852), 메모리(864), 디스플레이(854), 통신 인터페이스(866) 및 트랜스시버(868) 각각은 다양한 버스를 사용하여 상호 연결되고, 몇몇 컴포넌트들은 공통 마더보드 상에 또는 적절한 다른 방식으로 장착될 수 있다.
- [0133] 프로세서(852)는 메모리(864)에 저장된 명령어들을 포함하는 모바일 컴퓨팅 장치(850) 내의 명령어들을 실행할 수 있다. 프로세서(852)는 별개의 그리고 다수의 아날로그 및 디지털 프로세서들을 포함하는 칩들의 칩셋으로서 구현될 수 있다. 프로세서(852)는 예를 들어 사용자 인터페이스들, 모바일 컴퓨팅 장치(850)에 의해 실행되는 애플리케이션 및 모바일 컴퓨팅 장치(850)에 의한 무선 통신과 같은 모바일 컴퓨팅 장치(850)의 다른 컴포넌트들의 조정을 제공할 수 있다.
- [0134] 프로세서(852)는 제어 인터페이스(858) 및 디스플레이(856)에 연결된 디스플레이 인터페이스(856)를 통해 사용자와 통신할 수 있다. 디스플레이(854)는 예를 들어, TFT LCD(박막 트랜지스터 액정 디스플레이) 또는 OLED(유기 발광 다이오드) 디스플레이 또는 다른 적절한 디스플레이 기술을 포함할 수 있다. 디스플레이 인터페이스(856)는 사용자에게 그래픽 및 다른 정보를 제공하기 위해 디스플레이(854)를 구동하기 위한 적절한 회로를 포함할 수 있다. 제어 인터페이스(858)는 사용자로부터 명령을 수신하고, 프로세서(852)에 제출하기 위해 그들을 변환할 수 있다. 추가로, 외부 인터페이스(862)는 프로세서(852)와의 통신에 제공되어 다른 장치들과 모바일 컴퓨팅 장치(850)의 근거리 통신을 가능하게할 수 있다. 외부 인터페이스(862)는 예를 들면, 일부 구현예들에서는 유선 통신을 위해 또는 다른 구현예들에서는 무선 통신을 위해 제공될 수 있고, 다수의 인터페이스들도 사용될 수 있다.
- [0135] 메모리(864)는 모바일 컴퓨팅 장치(850) 내에 정보를 저장한다. 메모리(864)는 컴퓨터 판독가능 매체 또는 매체들, 휘발성 메모리 유닛 또는 유닛들, 비휘발성 메모리 유닛 또는 유닛들 중 하나 이상으로 구현될 수 있다. 또한 확장 메모리(874)는 예를 들어 SIMM(Single In Line Memory Module) 카드 인터페이스를 포함할 수 있는 확장 인터페이스(872)를 통해 모바일 컴퓨팅 장치(850)에 제공되고 접속될 수 있다. 상기 확장 메모리(874)는 모바일 컴퓨팅 장치(850)에 대해 추가 저장 공간을 제공할 수 있거나, 또는 애플리케이션들 또는 모바일 컴퓨팅 장치(850)에 대한 다른 정보를 저장할 수 있다. 특히, 확장 메모리(874)는 전술한 프로세스들을 수행하거나 보충하는 명령어들을 포함할 수 있으며, 또한 보안 정보를 포함할 수 있다. 따라서, 예를 들면, 확장 메모리(87

4)는 모바일 컴퓨팅 장치(850)에 대한 보안 모듈로서 제공될 수 있고, 모바일 컴퓨팅 장치(850)의 보안 사용을 허용하는 명령어들로 프로그래밍될 수 있다. 또한, 보안 애플리케이션들은 SIMM 카드 상의 식별 정보를 해킹할 수 없는 방식으로 배치하는 것과 같이, SIMM 카드들을 통해 추가 정보와 함께 제공될 수 있다.

[0136] 메모리는 예를 들어, 후술되는 바와 같이, 플래시 메모리 및/또는 NVRAM 메모리(non-volatile random access memory)를 포함할 수 있다. 일부 구현예들에서, 명령어들은 정보 캐리어에 저장되어, 하나 이상의 프로세싱 장치들(예를 들어, 프로세서(852))에 의해 실행될 때, 상술한 것과 같은 하나 이상의 방법들을 수행한다. 명령어들은 또한 하나 이상의 컴퓨터 또는 기계 판독가능 매체(예를 들어, 메모리(864), 확장 메모리(874) 또는 프로세서(852)상의 메모리)와 같은 하나 이상의 저장 장치들에 의해 저장될 수 있다. 일부 구현예들에서, 명령어들은 예를 들어 트랜스시버(868) 또는 외부 인터페이스(862)를 통해 전파된 신호로 수신될 수 있다.

[0137] 모바일 컴퓨팅 장치(850)는 필요에 따라 디지털 신호 처리 회로를 포함할 수 있는 통신 인터페이스(866)를 통해 무선으로 통신할 수 있다. 통신 인터페이스(866)는 GSM 음성 호출(이동 통신용 글로벌 시스템), SMS(단문 메시지 서비스), EMS(확장 메시징 서비스) 또는 MMS 메시징(멀티미디어 메시징 서비스), CDMA(코드 분할 다중 접속), TDMA(시분할 다중 접속), PDC(개인 디지털 셀룰러), WCDMA(광대역 코드 분할 다중 접속), CDMA2000 또는 GPRS(일반 패킷 무선 서비스)와 같은 다양한 모드 또는 프로토콜 하에서 통신을 제공할 수 있다. 이러한 통신은 예를 들어, 무선 주파수를 사용하는 트랜스시버(868)를 통해 발생될 수 있다. 추가로, 블루투스, Wi-Fi 또는 다른 트랜시버(도시되지 않음)를 사용하는 것과 같은 단거리 통신이 발생될 수 있다. 추가로, GPS(Global Positioning System) 수신기 모듈(870)은 모바일 컴퓨팅 장치(850)상에서 실행되는 애플리케이션들에 의해 적절히 사용될 수 있는 추가적인 내비게이션 및 위치 관련 무선 데이터를 모바일 컴퓨팅 장치(850)에 제공할 수 있다.

[0138] 모바일 컴퓨팅 장치(850)는 또한 사용자로부터 발화된 정보를 수신하고 그것을 이용가능한 디지털 정보로 변환할 수 있는 오디오 코덱(860)을 사용하여 청각적으로 통신할 수 있다. 오디오 코덱(860)은 마찬가지로, 예를 들어 모바일 컴퓨팅 장치(850)의 핸드셋 내의 스피커를 통하는 것과 같이, 사용자를 위한 가청 사운드를 생성할 수 있다. 이러한 사운드는 보이스 전화 콜들로부터의 사운드 포함할 수 있고, 기록된 사운드(예를 들어, 음성 메시지, 음악 파일 등)를 포함할 수 있고, 또한 모바일 컴퓨팅 장치(850)상에서 동작하는 애플리케이션들에 의해 생성된 사운드를 포함할 수 있다.

[0139] 모바일 컴퓨팅 장치(850)는 도면에 도시된 바와 같이 다수의 상이한 형태로 구현될 수 있다. 예를 들면, 그것은 셀룰러 전화(880)로서 구현될 수 있다. 또한 스마트폰(882), 개인 휴대 정보 단말기(PDA), 또는 다른 유사한 모바일 장치의 일부로서 구현될 수 있다.

[0140] 본 명세서에서 설명된 요지 및 기능적 동작의 실시 예는 디지털 전자 회로, 유형적으로 구현된 컴퓨터 소프트웨어 또는 펌웨어, 컴퓨터 하드웨어로 구현될 수 있으며, 본 명세서 및 그 구조적 균등물에 개시된 구조를 포함하거나, 그들 중 하나 이상의 조합을 포함한다. 본 명세서에서 설명된 요지의 실시 예는 하나 이상의 컴퓨터 프로그램, 즉 데이터 프로세싱 장치에 의해 실행되거나 또는 데이터 프로세싱 장치의 동작을 제어하기 위한 유형의 비-일시적인 프로그램 캐리어상에 인코딩된 컴퓨터 프로그램 명령어들의 하나 이상의 모듈로서 구현될 수 있다. 선택적으로 또는 부가적으로, 상기 프로그램 명령어들은 데이터 프로세싱 장치에 의한 실행을 위해 적절한 수신기 장치로의 송신을 위한 정보를 인코딩하기 위해 생성된 인위적으로 생성된 전파 신호, 예를 들어, 기계 - 생성 전기, 광학 또는 전자기 신호 상에 인코딩될 수 있다. 컴퓨터 저장 매체는 기계 판독 가능 저장 장치, 기계 판독 가능 저장 기관, 랜덤 또는 시리얼 액세스 메모리 장치, 또는 이들 중 하나 이상의 조합일 수 있다.

[0141] "데이터 프로세싱 장치"라는 용어는 예를 들어 프로그램 가능 프로세서, 컴퓨터, 또는 복수의 프로세서 또는 컴퓨터를 포함하는 데이터 프로세싱을 위한 모든 종류의 장치, 디바이스 및 기계를 포함한다. 상기 장치는 특수 목적 논리 회로, 예를 들어 FPGA(field programmable gate array) 또는 ASIC(application specific integrated circuit)을 포함할 수 있다. 또한, 상기 장치는 하드웨어 이외에, 해당 컴퓨터 프로그램에 대한 실행 환경을 생성하는 코드, 예를 들어 프로세서 펌웨어, 프로토콜 스택, 데이터베이스 관리 시스템, 운영 체제, 또는 이들 중 하나 이상의 조합을 구성하는 코드를 포함할 수 있다.

[0142] 컴퓨터 프로그램(프로그램, 소프트웨어, 소프트웨어 애플리케이션, 모듈, 소프트웨어 모듈, 스크립트 또는 코드로 지칭되거나 설명될 수 있음)은 컴파일된 또는 해석된 언어, 또는 선언적 또는 절차적 언어를 포함한 모든 형태의 프로그래밍 언어로 작성될 수 있으며, 독립 실행형 프로그램이나 모듈, 컴포넌트, 서브루틴 또는 컴퓨팅 환경에서 사용하기에 적합한 다른 장치를 포함하여 어떤 형태로든 배포될 수 있다. 컴퓨터 프로그램은 파일 시스템의 파일에 해당할 수 있지만 반드시 그런 것은 아니다. 프로그램은 프로그램 전용 단일 파일, 여러 개의 조

정된 파일(예를 들어, 하나 이상의 모듈, 하위 프로그램 또는 코드의 부분들을 저장하는 파일들), 또는 마크업 언어 문서에 저장된 하나 이상의 스크립트와 같은 다른 프로그램들 또는 데이터를 보유하고 있는 파일의 부분에 저장될 수 있다. 컴퓨터 프로그램은 한 사이트에 있거나 여러 사이트에 분산되어 있으며 통신 네트워크로 상호 연결된 여러 대의 컴퓨터 또는 하나의 컴퓨터에서 실행되도록 배포될 수 있다.

[0143] 본 명세서에서 설명되는 프로세스들 및 로직 흐름은 입력 데이터를 오퍼레이팅하고 출력을 생성함으로써 기능들을 수행하기 위해 하나 이상의 컴퓨터 프로그램을 실행하는 하나 이상의 프로그램 가능 컴퓨터에 의해 수행될 수 있다. 상기 프로세스들 및 로직 흐름은 또한 FPGA(field programmable gate array) 또는 ASIC(application-specific integrated circuit)과 같은 특수 목적의 논리 회로에 의해 수행될 수 있고, 장치는 또한 상기 특수 목적의 논리 회로로 구현될 수 있다.

[0144] 컴퓨터 프로그램의 실행에 적합한 컴퓨터들은 예를 들어 범용 또는 특수 목적 마이크로프로세서들 또는 둘 모두, 또는 임의의 다른 종류의 중앙 처리 장치를 포함하고, 이들에 기반될 수 있다. 일반적으로, 중앙 처리 장치는 판독 전용 메모리 또는 랜덤 액세스 메모리 또는 둘 모두로부터 명령어들과 데이터를 수신할 것이다. 컴퓨터의 필수 구성요소들은 명령어들을 수행하거나 실행하기 위한 중앙 처리 장치 및 명령어들과 데이터를 저장하기 위한 하나 이상의 메모리 장치이다. 일반적으로, 컴퓨터는 예를 들어, 자기, 광 자기 디스크 또는 광학 디스크와 같은 데이터를 저장하기 위한 하나 이상의 대용량 저장 장치로부터 데이터를 수신하거나 그 하나 이상의 대용량 저장 장치에 데이터를 전송하기 위해 동작 가능하게 결합될 것이다. 그러나, 컴퓨터는 그러한 장치들을 가질 필요가 없다. 또한, 컴퓨터는 다른 장치, 예를 들어, 이동 전화기, 개인 휴대 정보 단말기(PDA), 이동 오디오 또는 비디오 플레이어, 게임 콘솔, GPS 수신기 또는 예를 들어 범용 직렬 버스(USB) 플래시 드라이브와 같은 휴대용 저장 장치에 내장될 수 있다.

[0145] 컴퓨터 프로그램 명령어들과 데이터를 저장하기에 적합한 컴퓨터 판독 가능 매체는 예를 들어 EPROM, EEPROM 및 플래시 메모리 장치들과 같은 반도체 메모리 장치들; 내부 하드 디스크 또는 이동식 디스크와 같은 자기 디스크; 광 자기 디스크; 그리고 CD-ROM 및 DVD-ROM 디스크를 포함하는 모든 형태의 비휘발성 메모리, 매체 및 메모리 장치들을 포함한다. 상기 프로세서 및 메모리는 특수 목적 논리 회로에 의해 통합되거나 보완될 수 있다.

[0146] 사용자와의 상호 작용을 제공하기 위해, 본 명세서에서 설명된 요지의 실시예들은 사용자에게 정보를 디스플레이하기 위한 디스플레이 장치(예를 들어 CRT(cathode ray tube) 모니터, LCD(liquid crystal display) 모니터), 사용자가 컴퓨터에 입력을 제공할 수 있는 키보드 및 포인팅 장치(예를 들어 마우스 또는 트랙볼)를 갖는 컴퓨터상에서 구현될 수 있다. 다른 종류의 디바이스들이 사용자와의 상호 작용을 제공하는 데 사용될 수 있으며, 예를 들어, 사용자에게 제공되는 피드백은 시각 피드백, 청각 피드백 또는 촉각 피드백과 같은 임의의 형태의 감각 피드백일 수 있으며, 사용자로부터의 입력은 음향, 음성 또는 촉각 입력을 포함하는 임의의 형태로 수신될 수 있다. 또한, 컴퓨터는 사용자가 사용하는 디바이스로부터 자원들을 수신하고 그 자원들을 보냄으로써, 예를 들어, 웹 브라우저로부터 수신된 요청들에 응답하여 사용자의 클라이언트 디바이스상의 웹 브라우저에 웹 페이지들을 전송함으로써 사용자와 상호 작용할 수 있다.

[0147] 본 명세서에서 설명된 요지의 실시예들은 데이터 서버와 같은 백 엔드 컴포넌트, 또는 예를 들어 애플리케이션 서버와 같은 미들웨어 컴포넌트, 또는 본 명세서에 설명된 요지의 구현예와 사용자가 상호 작용할 수 있는 웹 브라우저를 또는 그래픽 사용자 인터페이스를 갖는 클라이언트 컴퓨터와 같은 프론트 엔드 컴포넌트, 또는 하나 이상의 백엔드, 미들웨어 또는 프론트 엔드 컴포넌트의 모든 조합을 포함하는 컴퓨팅 시스템으로 구현될 수 있다. 상기 시스템의 컴포넌트들은 디지털 데이터 통신의 임의의 형태 또는 매체, 예를 들어 통신 네트워크에 의해 상호 접속될 수 있다. 예시적인 통신 네트워크들은 근거리 통신망("LAN") 및 광역 통신망("WAN"), 예를 들어 인터넷을 포함한다.

[0148] 상기 컴퓨팅 시스템은 클라이언트들과 서버들을 포함할 수 있다. 클라이언트와 서버는 일반적으로 서로 멀리 떨어져 있으며, 일반적으로 통신 네트워크를 통해 상호 작용한다. 클라이언트와 서버의 관계는 각각의 컴퓨터들에서 실행되고 서로 클라이언트-서버 관계를 갖는 컴퓨터 프로그램들에 의해 발생한다.

[0149] 본 명세서는 다수의 특정 구현 세부 사항을 포함하지만, 이들은 청구될 수 있는 범위에 대한 제한으로서 해석되어서는 안되며, 오히려 특정 실시예들에 특정 될 수 있는 특징들에 대한 설명으로 해석되어야 한다. 별도의 실시예들과 관련하여 본 명세서에서 설명되는 특정 특징들은 또한 단일 실시예에서 조합하여 구현될 수 있다. 반대로, 단일 실시예의 문맥에서 설명된 다양한 특징은 또한 다수의 실시예에서 개별적으로 또는 임의의 적합한 하위 조합으로 구현될 수 있다. 더욱이, 특징들은 특정 조합으로 동작하는 것으로 상술될 수 있지만, 청구된 조합으로부터의 하나 이상의 특징이 일부 경우에서 그 조합으로부터 제거될 수 있으며, 청구된 조합은 서브 조합

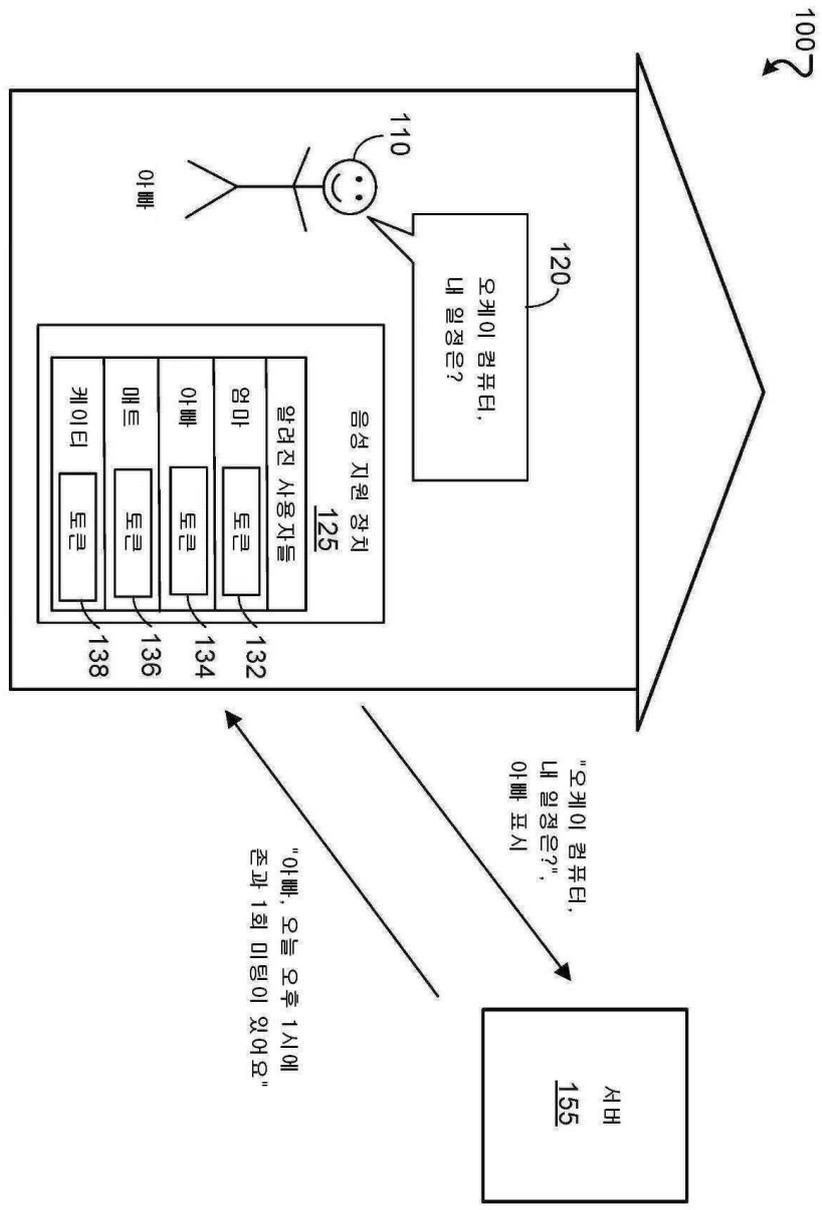
또는 서버 조합의 변형에 관한 것일 수 있다.

[0150] 유사하게, 동작들이 특정 순서로 도면들에 도시되어 있지만, 이는 바람직한 결과들을 달성하기 위해 상기 동작들이 도시된 특정 순서 또는 순차적 순서로 수행되거나 모든 예시된 동작이 수행될 것을 요구하는 것으로 이해되어서는 안된다. 특정 상황에서 멀티 태스킹 및 병렬 처리가 유리할 수 있다. 또한, 전술한 실시예들에서 다양한 시스템 컴포넌트들의 분리는 모든 실시예들에서 그러한 분리를 요구하는 것으로 이해되어서는 안되며, 서술된 프로그램 컴포넌트들 및 시스템들은 일반적으로 단일 소프트웨어 제품에 함께 통합되거나 다중 소프트웨어 제품들로 패키징될 수 있다는 것을 이해해야 한다.

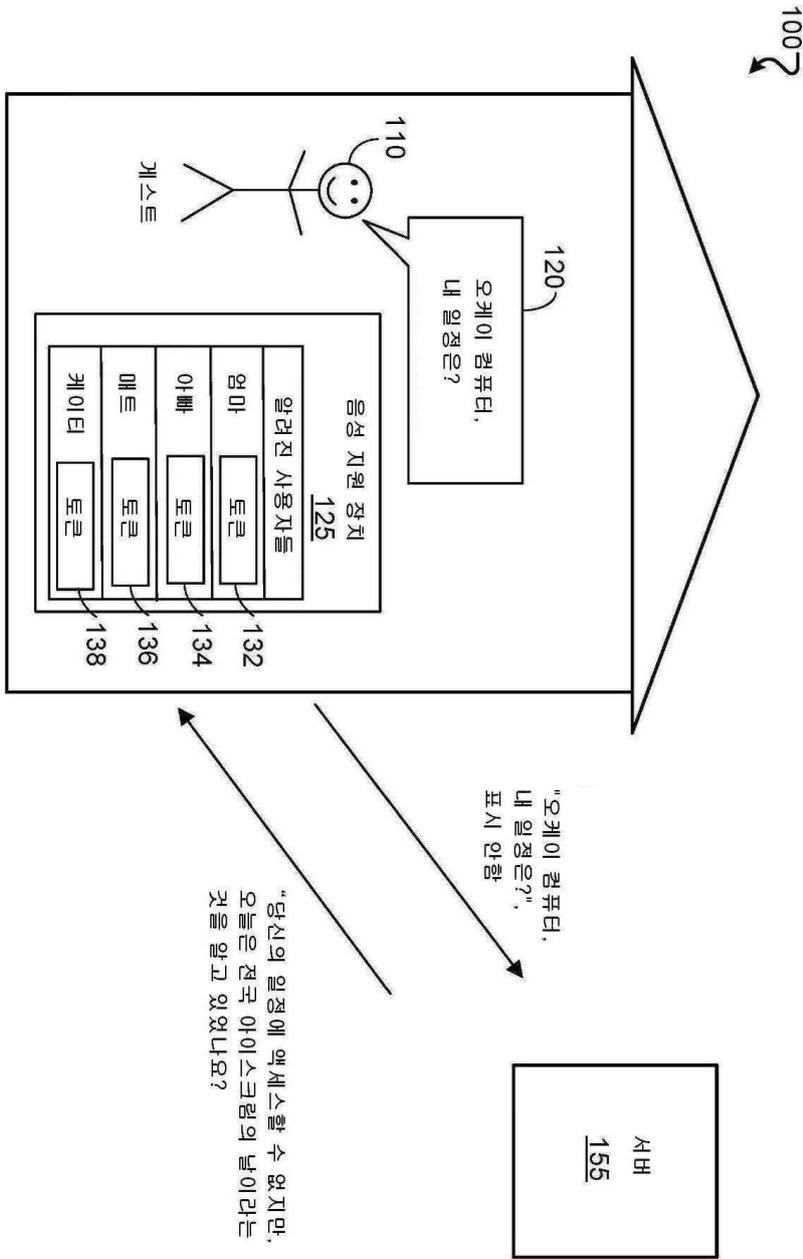
[0151] 본 발명의 특정 실시예들이 설명되었다. 다른 실시예들은 다음의 청구항들의 범위 내에 있다. 예를 들어, 청구 범위에 열거된 동작들은 상이한 순서로 수행될 수 있으며 여전히 바람직한 결과들을 달성한다. 하나의 예로서, 첨부된 도면들에 도시된 프로세스들은 바람직한 결과들을 얻기 위해 나타낸 특정 순서 또는 순차적 순서를 반드시 필요로 하지는 않는다. 특정 구현예들에서, 멀티태스킹 및 병렬 처리가 유리할 수 있다. 기술된 프로세스들로부터 다른 단계들이 제공되거나 단계들이 제거될 수 있다. 따라서, 다른 구현예들은 다음의 청구 범위의 범위 내에 있다.

**도면**

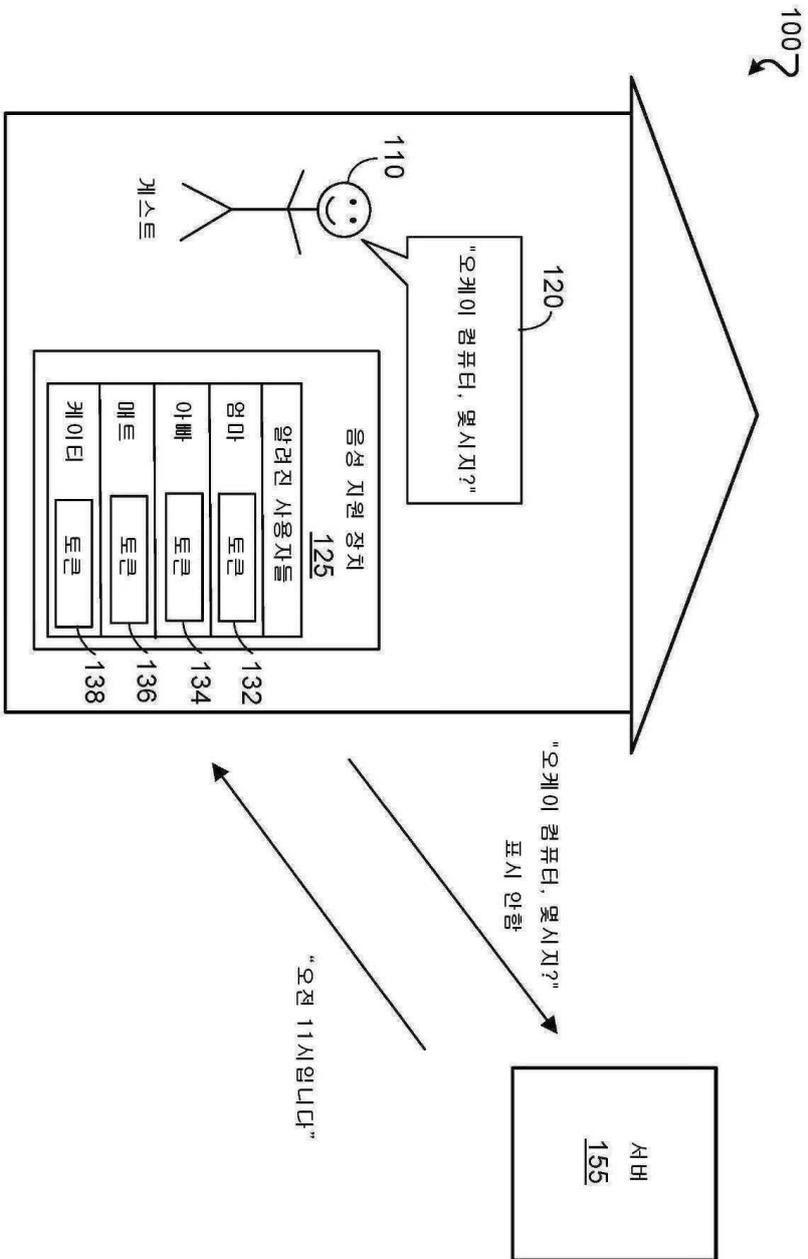
**도면1a**



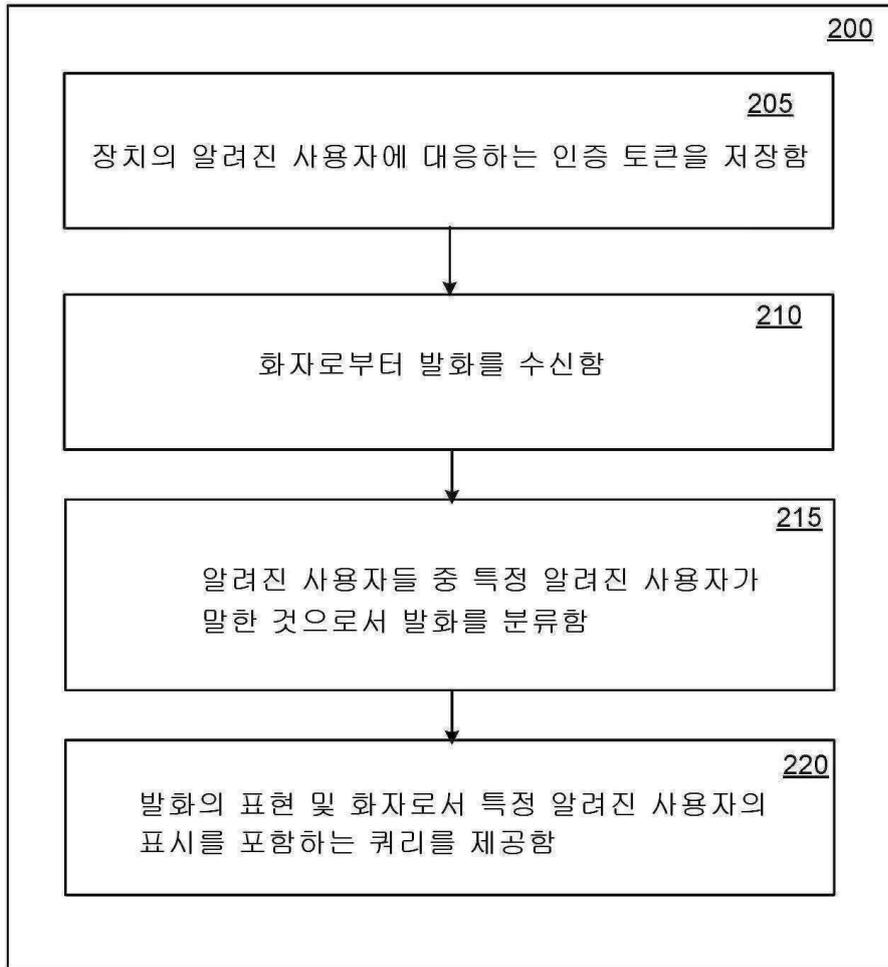
도면1b



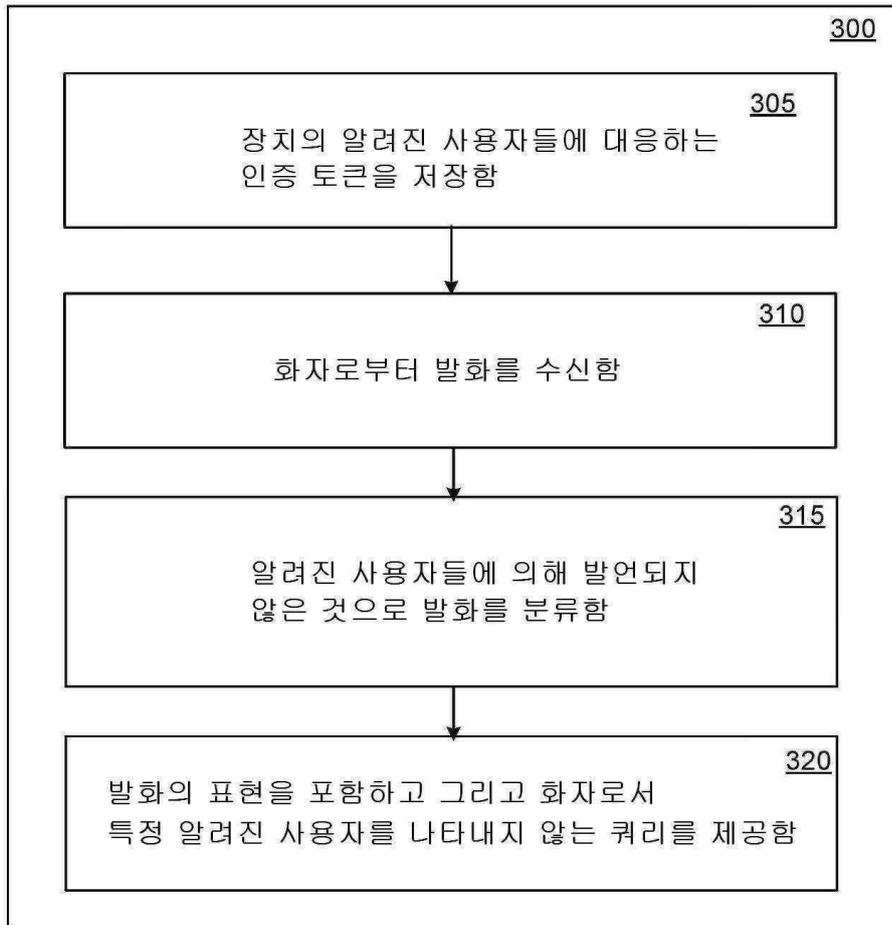
도면1c



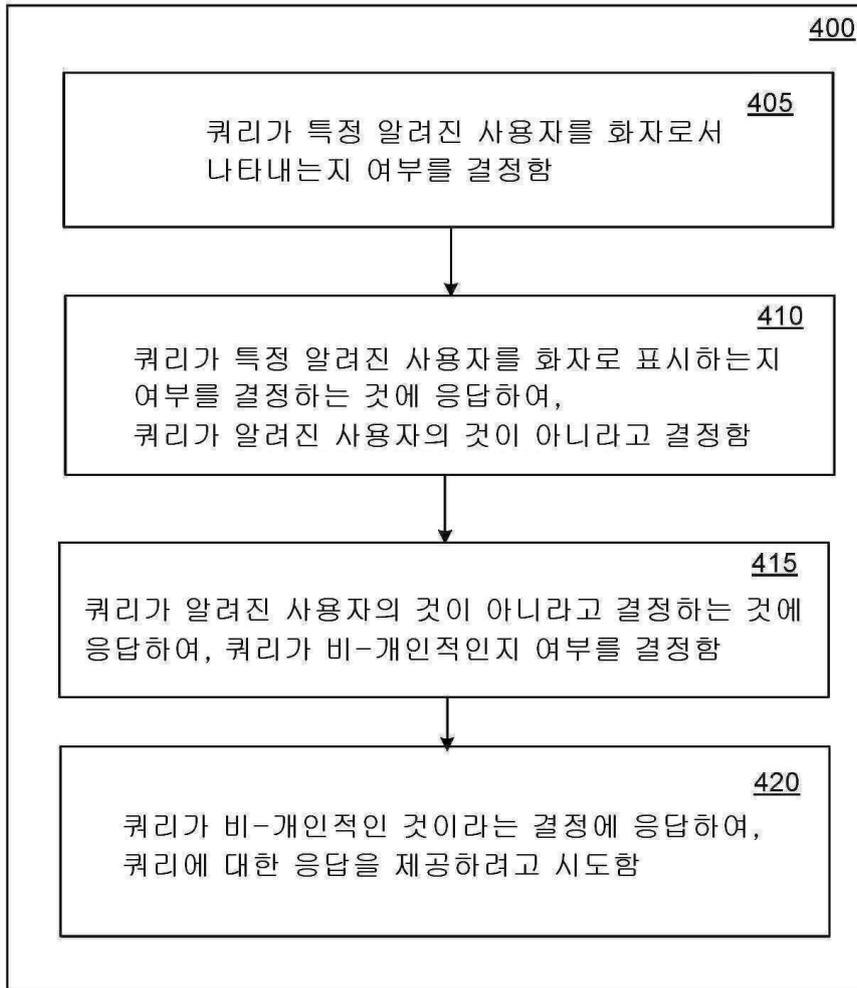
도면2



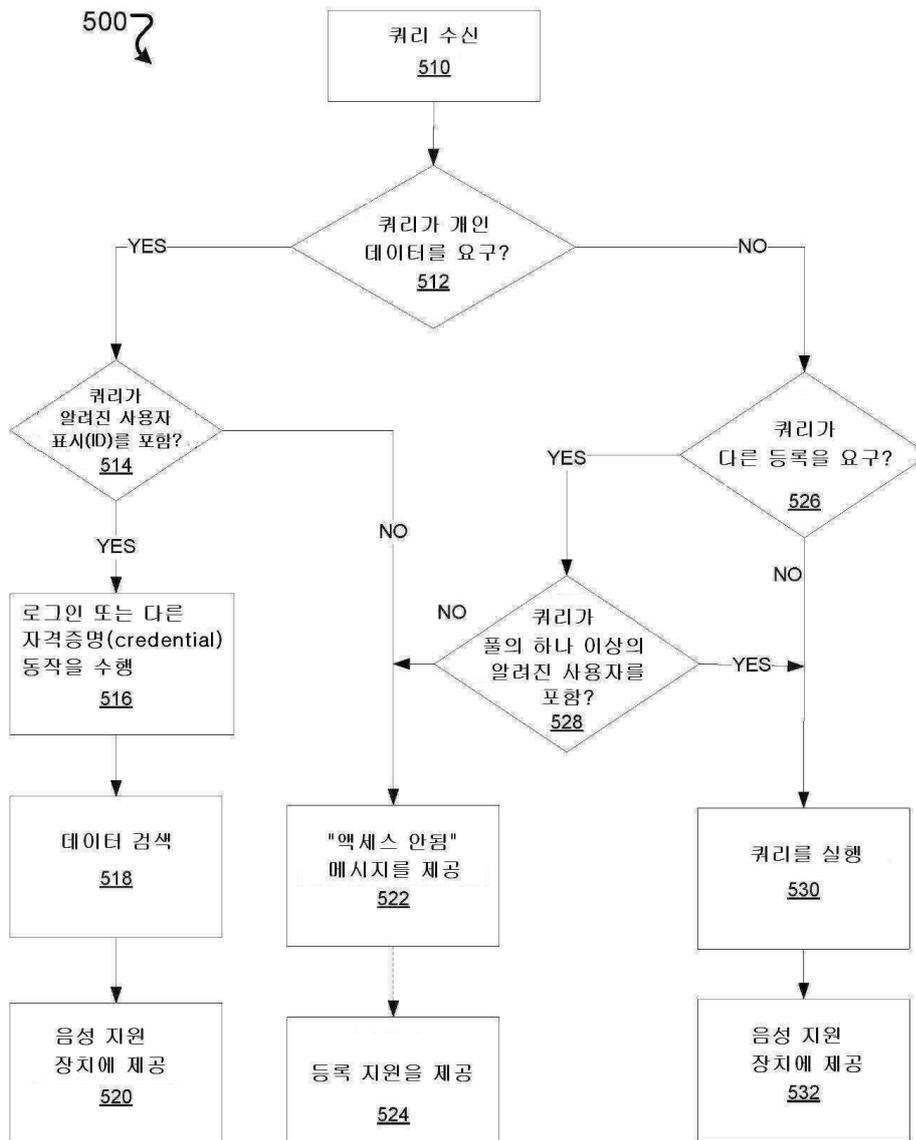
도면3



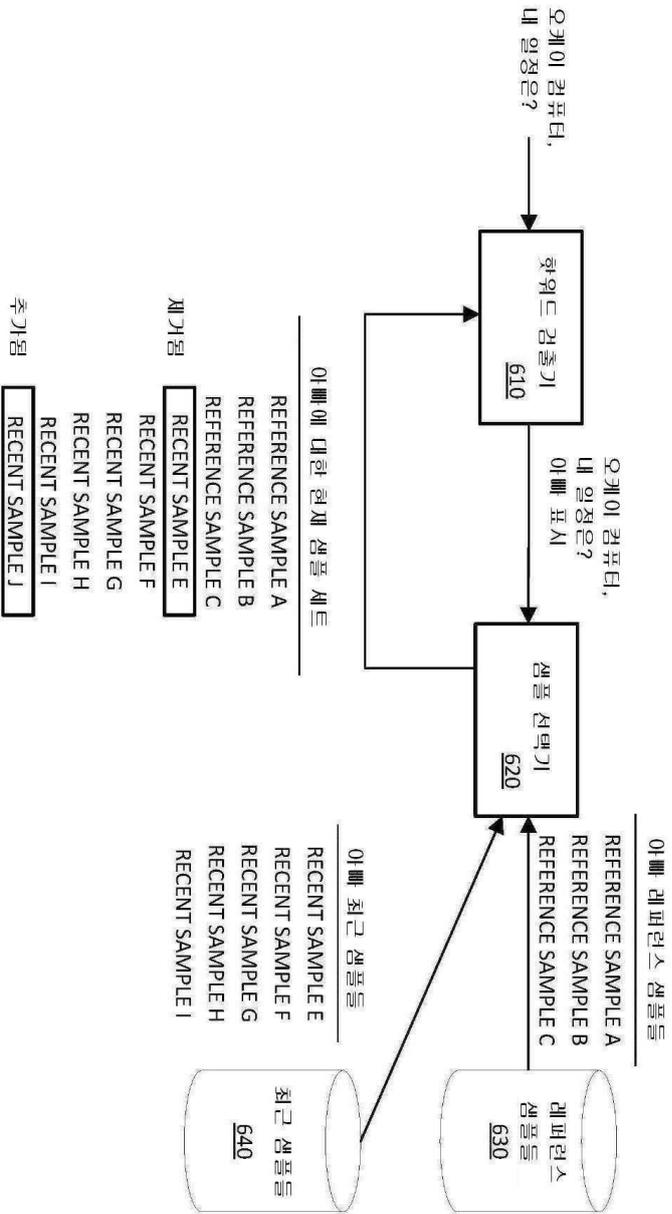
도면4



도면5

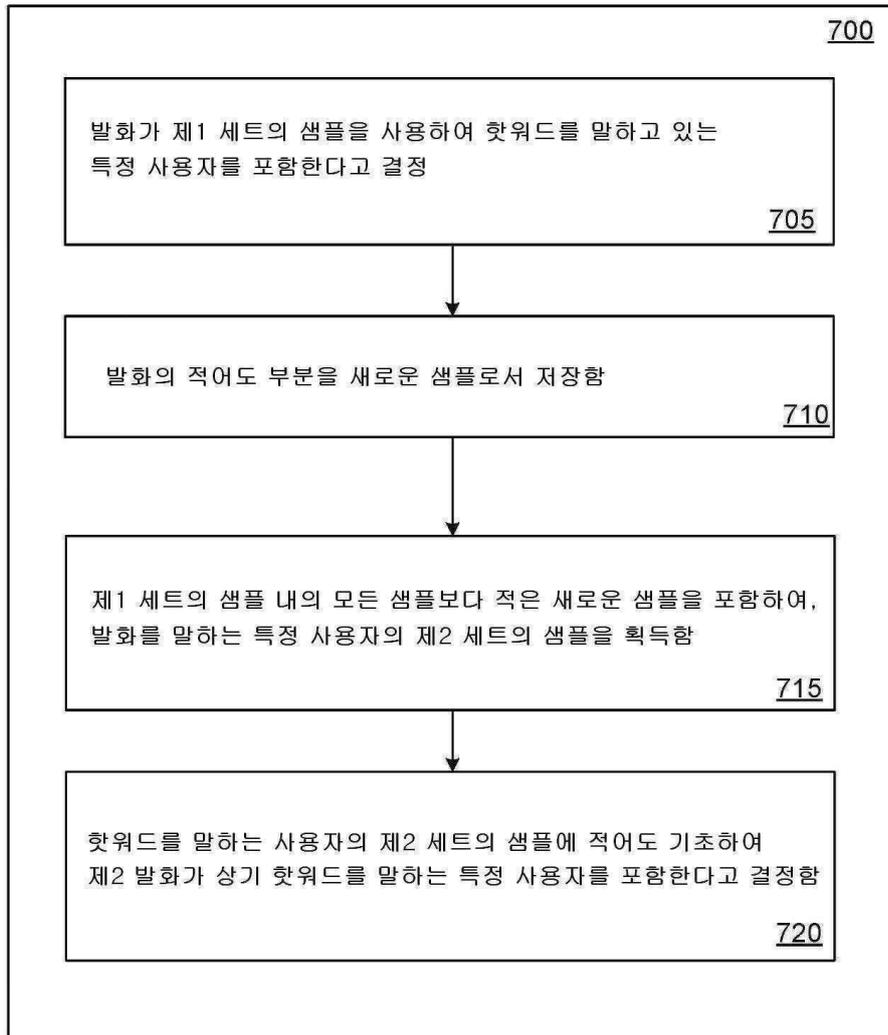


6002



도면6

도면7



도면8

