

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2019년 8월 29일 (29.08.2019) **WIPO | PCT**



(10) 국제공개번호
WO 2019/164120 A1

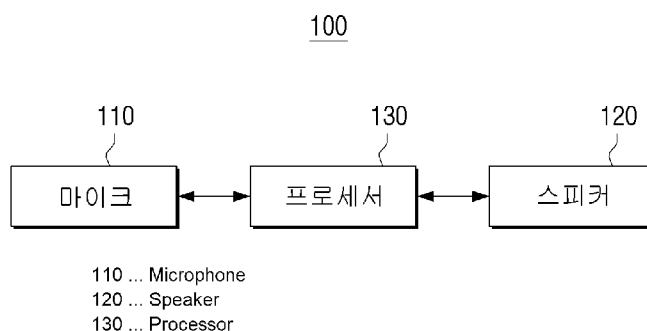
- (51) 국제특허분류:
G06F 3/16 (2006.01) *G06F 3/01* (2006.01)
G06F 3/00 (2006.01) *G06K 9/00* (2006.01)
- (21) 국제출원번호: PCT/KR2019/000036
- (22) 국제출원일: 2019년 1월 2일 (02.01.2019)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2018-0021876 2018년 2월 23일 (23.02.2018) KR
- (71) 출원인: 삼성전자주식회사 (**SAMSUNG ELECTRONICS CO., LTD.**) [KR/KR]; 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).
- (72) 발명자: 최형탁 (**CHOI, Hyungtak**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 고현목 (**KO, Hyeonmok**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 김지희 (**KIM, Jihie**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 김홍철 (**KIM, Hongchul**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR). 황인철 (**HWANG, Inchul**); 16677 경기도 수원시 영통구 삼성로 129, Gyeonggi-do (KR).
- (74) 대리인: 김태현 등 (**KIM, Tae-hun et al.**); 06626 서울특별시 서초구 강남대로 343 신덕빌딩 9층, Seoul (KR).

- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:
— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: ELECTRONIC DEVICE AND CONTROL METHOD THEREOF

(54) 발명의 명칭: 전자 장치 및 그 제어 방법



(57) Abstract: Disclosed is an electronic device. The electronic device comprises: a microphone; a speaker; and a processor electrically connected to the microphone and speaker, wherein the processor, when a first user's voice is input through the microphone, identifies a user who uttered the first user's voice and provides a first response sound, which is obtained by inputting the first user's voice to an artificial intelligence model learned through an artificial intelligence algorithm, through the speaker, and when a second user's voice is input through the microphone, identifies a user who uttered the second user's voice, and if the user who uttered the first user's voice is the same as the user who uttered the second user's voice, provides a second response sound, which is obtained by inputting the second user's voice and utterance history information to the artificial intelligence model, through the speaker. In particular, at least some of the methods of providing a response sound to a user's voice may use an artificial intelligence model learned in accordance with at least one of a machine learning, neural network, or deep learning algorithm.

(57) 요약서: 전자 장치가 개시된다. 본 전자 장치는 마이크, 스피커 및 마이크, 스피커와 전기적으로 연결된 프로세서를 포함하며, 프로세서는 마이크를 통해 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 사용자를 식별하고 인공지능 알고리즘을 통해 학습된 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 스피커를 통해 제공하며, 마이크를 통해 제2 사용자 음성이 입력되면, 제2 사용자 음성을 발화한 사용자를 식별하고, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 스피커를 통해 제공할 수 있다. 특히, 사용자 음성에 대한 응답 사운드를 제공하는 방법의 적어도 일부는 기계학습, 신경망 또는 딥러닝 알고리즘 중 적어도 하나에 따라 학습된 인공지능 모델이 이용될 수 있다.

명세서

발명의 명칭: 전자 장치 및 그 제어 방법

기술분야

- [1] 본 개시는 전자 장치 및 그 제어 방법에 대한 것으로, 더욱 상세하게는 사용자 음성을 인식하고, 인식된 사용자 음성에 따라 다양한 인터랙션을 제공하는 전자 장치 및 그 제어 방법에 대한 것이다.
- [2] 또한, 본 개시는 기계 학습 알고리즘을 활용하여 인간 두뇌의 인지, 판단 등의 기능을 모사하는 인공 지능(Artificial Intelligence, AI) 시스템 및 그 응용에 관한 것이다.

배경기술

- [3] 근래에는 인간 수준의 지능을 구현하는 인공 지능 시스템이 다양한 분야에서 이용되고 있다. 인공 지능 시스템은 기존의 룰(rule) 기반 스마트 시스템과 달리 기계가 스스로 학습하고 판단하며 똑똑해지는 시스템이다. 인공 지능 시스템은 사용할수록 인식률이 향상되고 사용자 취향을 보다 정확하게 이해할 수 있게 되어, 기존 룰 기반 스마트 시스템은 점차 딥러닝 기반 인공 지능 시스템으로 대체되고 있다.
- [4] 인공 지능 기술은 기계학습(예로, 딥러닝) 및 기계학습을 활용한 요소 기술들로 구성된다.
- [5] 기계학습은 입력 데이터들의 특징을 스스로 분류/학습하는 알고리즘 기술이며, 요소기술은 딥러닝 등의 기계학습 알고리즘을 활용하여 인간 두뇌의 인지, 판단 등의 기능을 모사하는 기술로서, 언어적 이해, 시각적 이해, 추론/예측, 지식 표현, 동작 제어 등의 기술 분야로 구성된다.
- [6] 인공 지능 기술이 응용되는 다양한 분야는 다음과 같다. 언어적 이해는 인간의 언어/문자를 인식하고 응용/처리하는 기술로서, 자연어 처리, 기계 번역, 대화시스템, 질의 응답, 음성 인식/합성 등을 포함한다. 시각적 이해는 사물을 인간의 시각처럼 인식하여 처리하는 기술로서, 객체 인식, 객체 추적, 영상 검색, 사람 인식, 장면 이해, 공간 이해, 영상 개선 등을 포함한다. 추론 예측은 정보를 판단하여 논리적으로 추론하고 예측하는 기술로서, 지식/확률 기반 추론, 최적화 예측, 선호 기반 계획, 추천 등을 포함한다. 지식 표현은 인간의 경험정보를 지식데이터로 자동화 처리하는 기술로서, 지식 구축(데이터 생성/분류), 지식 관리(데이터 활용) 등을 포함한다. 동작 제어는 차량의 자율 주행, 로봇의 움직임을 제어하는 기술로서, 움직임 제어(항법, 충돌, 주행), 조작 제어(행동 제어) 등을 포함한다.
- [7] 한편, 종래에는 사용자 음성을 인식하고, 인식된 사용자 음성에 따라 다양한 인터랙션을 제공하는 전자 장치들이 개발되고 있다. 또한, 전자 장치들은 일회적인 사용자 음성에 반응한 응답을 제공할 뿐만 아니라, 연속적인 사용자

음성을 수신하고, 발화 히스토리를 반영하여 응답을 제공할 수도 있다. 이때, 연속적인 사용자 음성 사이에 다른 사용자의 사용자 음성이 수신되면 이를 구분하지 못하고, 전혀 별개의 응답을 제공하는 문제가 있었다.

발명의 상세한 설명

기술적 과제

- [8] 본 개시는 상술한 필요성에 따른 것으로, 본 개시의 목적은 복수의 사용자 각각으로부터 사용자 음성이 입력되더라도 컨텍스트(context)가 유지되는 응답을 제공하는 전자 장치 및 그 제어 방법을 제공함에 있다.

과제 해결 수단

- [9] 이상과 같은 목적을 달성하기 위한 본 개시의 일 실시 예에 따르면, 전자 장치는 마이크, 스피커 및 상기 마이크 및 상기 스피커와 전기적으로 연결된 프로세서를 포함하며, 상기 프로세서는 상기 마이크를 통해 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 사용자를 식별하고 인공지능 알고리즘을 통해 학습된 인공지능 모델에 상기 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 상기 스피커를 통해 제공하며, 상기 마이크를 통해 제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을 발화한 사용자를 식별하고, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 상기 스피커를 통해 제공할 수 있다.
- [10] 또한, 스토리지를 더 포함하고, 상기 프로세서는 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 상기 제2 사용자 음성을 상기 스토리지에 저장하고, 상기 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자 음성이 입력되지 않으면, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 제공할 수 있다.
- [11] 그리고, 상기 프로세서는 상기 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 제1 사용자를 식별하고 상기 인공지능 모델에 상기 제1 사용자 음성을 입력하여 획득된 상기 제1 응답 사운드를 상기 스피커를 통해 제공하며, 상기 제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을 발화한 제2 사용자를 식별하고, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 상기 제1 응답 사운드와 구분되게 제공할 수 있다.
- [12] 또한, 상기 프로세서는 상기 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 상기 인공지능 모델에 상기 제1 추가 사용자 음성 및 상기 제1 사용자 음성에 대응되는 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를 상기 스피커를 통해 제공하고, 상기 제2 사용자에 의해 제2 추가 사용자 음성이 입력되면, 상기 인공지능 모델에 상기 제2 추가 사용자 음성 및

상기 제2 사용자 음성에 대응되는 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 상기 스피커를 통해 상기 제1 추가 응답 사운드와 구분되게 제공할 수 있다.

- [13] 그리고, 상기 프로세서와 전기적으로 연결된 디스플레이를 더 포함하고, 상기 프로세서는 상기 제1 추가 응답 사운드를 제공하며 상기 제1 사용자에게 대응되는 제1 UI를 상기 디스플레이를 통해 디스플레이하고, 상기 제2 추가 응답 사운드를 제공하며 상기 제2 사용자에게 대응되는 제2 UI를 상기 디스플레이를 통해 디스플레이할 수 있다.
- [14] 또한, 상기 프로세서는 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인을 획득하고, 상기 제1 도메인 및 상기 제2 도메인이 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 상기 발화 히스토리 정보를 입력하여 획득된 상기 제2 응답 사운드를 상기 스피커를 통해 제공할 수 있다.
- [15] 그리고, 상기 프로세서와 전기적으로 연결된 스토리지를 더 포함하고, 상기 프로세서는 상기 제2 사용자 음성이 상기 제1 사용자 음성이 입력된 시점으로부터 기설정된 제1 시간 내에 입력되고, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하며, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인이 서로 다르면, 상기 제1 응답 사운드의 제공을 생략하고 상기 제1 사용자 음성을 상기 스토리지에 저장하며, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 제공할 수 있다.
- [16] 또한, 상기 프로세서는 상기 사용자로부터 상기 제2 도메인에 대응되는 사용자 음성이 기설정된 제2 시간 이상 입력되지 않으면, 상기 인공지능 모델에 상기 스토리지에 저장된 상기 제1 사용자 음성을 입력하여 획득된 상기 제1 응답 사운드를 상기 스피커를 통해 제공할 수 있다.
- [17] 그리고, 카메라를 더 포함하고, 상기 프로세서는 상기 카메라에 의해 촬영된 영상 및 상기 마이크를 통해 입력된 사용자 음성 중 적어도 하나에 기초하여 상기 사용자를 식별할 수 있다.
- [18] 또한, 상기 프로세서는 상기 제1 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된 사용자의 입모양에 기초하여 상기 제1 사용자 음성을 발화한 사용자를 식별하고, 상기 제2 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된 상기 식별된 사용자의 입모양이 변경되면, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일한 것으로 식별할 수 있다.
- [19] 한편, 본 개시의 일 실시 예에 따르면, 전자 장치의 제어 방법은 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 사용자를 식별하는 단계, 인공지능 알고리즘을 통해 학습된 인공지능 모델에 상기 제1 사용자 음성을

- 입력하여 획득된 제1 응답 사운드를 제공하는 단계, 제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을 발화한 사용자를 식별하는 단계 및 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공하는 단계를 포함할 수 있다.
- [20] 또한, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 상기 제2 사용자 음성을 저장하는 단계 및 상기 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자 음성이 입력되지 않으면, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 제공하는 단계를 더 포함할 수 있다.
- [21] 그리고, 상기 제1 사용자 음성을 발화한 사용자를 식별하는 단계는 상기 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 제1 사용자를 식별하고, 상기 제1 응답 사운드를 제공하는 단계는 상기 인공지능 모델에 상기 제1 사용자 음성을 입력하여 획득된 상기 제1 응답 사운드를 제공하며, 상기 제2 사용자 음성을 발화한 사용자를 식별하는 단계는 상기 제2 사용자 음성을 발화한 제2 사용자를 식별하고, 상기 제어 방법은 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 상기 제1 응답 사운드와 구분되게 제공하는 단계를 더 포함할 수 있다.
- [22] 또한, 상기 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 상기 인공지능 모델에 상기 제1 추가 사용자 음성 및 상기 제1 사용자 음성에 대응되는 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를 제공하는 단계 및 상기 제2 사용자에 의해 제2 추가 사용자 음성이 입력되면, 상기 인공지능 모델에 상기 제2 추가 사용자 음성 및 상기 제2 사용자 음성에 대응되는 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 상기 제1 추가 응답 사운드와 구분되게 제공하는 단계를 더 포함할 수 있다.
- [23] 그리고, 상기 제1 추가 응답 사운드를 제공하는 단계는 상기 제1 추가 응답 사운드를 제공하며 상기 제1 사용자에 대응되는 제1 UI를 디스플레이하고, 상기 제2 추가 응답 사운드를 상기 제1 추가 응답 사운드와 구분되게 제공하는 단계는 상기 제2 추가 응답 사운드를 제공하며 상기 제2 사용자에 대응되는 제2 UI를 디스플레이할 수 있다.
- [24] 또한, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인을 획득하는 단계 및 상기 제1 도메인 및 상기 제2 도메인이 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 상기 발화 히스토리 정보를 입력하여 획득된 상기 제2 응답 사운드를 제공하는 단계를 더 포함할 수 있다.
- [25] 그리고, 상기 제어 방법은 상기 제2 사용자 음성이 상기 제1 사용자 음성이 입력된 시점으로부터 기설정된 제1 시간 내에 입력되고, 상기 제1 사용자 음성을

발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하며, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인이 서로 다르면, 상기 제1 응답 사운드의 제공을 생략하고 상기 제1 사용자 음성을 저장하며, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 제공할 수 있다.

[26] 또한, 상기 사용자로부터 상기 제2 도메인에 대응되는 사용자 음성이 기설정된 제2 시간 이상 입력되지 않으면, 상기 인공지능 모델에 상기 제1 사용자 음성을 입력하여 획득된 상기 제1 응답 사운드를 제공하는 단계를 더 포함할 수 있다.

[27] 그리고, 상기 사용자를 식별하는 단계는 촬영된 영상 및 상기 입력된 사용자 음성 중 적어도 하나에 기초하여 상기 사용자를 식별할 수 있다.

[28] 또한, 상기 제1 사용자 음성을 발화한 사용자를 식별하는 단계는 상기 제1 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된 사용자의 입모양에 기초하여 상기 제1 사용자 음성을 발화한 사용자를 식별하고, 상기 제2 응답 사운드를 제공하는 단계는 상기 제2 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된 상기 식별된 사용자의 입모양이 변경되면, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일한 것으로 식별할 수 있다.

발명의 효과

[29] 이상과 같은 본 개시의 다양한 실시 예에 따르면, 전자 장치는 복수의 사용자 각각으로부터 사용자 음성이 입력되면, 기존 사용자 음성에 기초하여 복수의 사용자 음성을 처리, 유보 또는 별도 처리함에 따라 컨텍스트를 유지한 채로 응답을 제공할 수 있다.

도면의 간단한 설명

[30] 도 1a 및 도 1b는 본 개시의 일 실시 예에 따른 복수의 사용자 음성을 처리하는 방법을 설명하기 위한 도면들이다.

[31] 도 2는 본 개시의 일 실시 예에 따른 사용자 음성에 대한 응답 사운드를 제공하는 전자 시스템을 도시한 도면이다.

[32] 도 3a는 본 개시의 일 실시 예에 따른 전자 장치의 구성을 설명하기 위한 블록도이다.

[33] 도 3b는 본 개시의 일 실시 예에 따른 전자 장치의 세부 구성을 나타내는 블록도이다.

[34] 도 4a 및 도 4b는 본 개시의 일 실시 예에 따른 복수의 사용자 중 한 명의 사용자 음성 처리의 연기를 설명하기 위한 도면들이다.

[35] 도 5a 내지 도 5c는 본 개시의 일 실시 예에 따른 복수의 사용자 음성의 실시간 동시 처리 방법을 설명하기 위한 도면들이다.

[36] 도 6은 본 개시의 일 실시 예에 따른 복수의 사용자와의 통합 인터랙션을 설명하기 위한 도면이다.

- [37] 도 7은 본 개시의 일 실시 예에 따른 동일 사용자의 복수의 도메인 각각에 대응하는 복수의 사용자 음성이 입력된 경우의 동작을 설명하기 위한 도면이다.
- [38] 도 8a 및 도 8b는 본 개시의 일 실시 예에 따른 사용자 음성을 발화한 사용자를 식별하기 위한 방법을 설명하기 위한 도면들이다.
- [39] 도 9a 및 도 9b는 본 개시의 일 실시 예에 따른 인공지능 모델의 추가 학습을 설명하기 위한 도면들이다.
- [40] 도 10은 본 개시의 일 실시 예에 따른 인공지능 모델이 저장된 전자 장치의 동작을 설명하기 위한 흐름도이다.
- [41] 도 11은 본 개시의 일 실시 예에 따른 전자 장치 및 인공지능 모델이 저장된 외부 서버의 동작을 설명하기 위한 시퀀스도이다.
- [42] 도 12는 본 개시의 일 실시 예에 따른 사용자 문의에 대한 응답을 제공하는 인공지능 모델을 학습하고 이용하기 위한 외부 서버의 구성을 나타내는 블록도이다.
- [43] 도 13은 본 개시의 일 실시 예에 따른 학습부를 나타내는 블록도이다.
- [44] 도 14는 본 개시의 일 실시 예에 따른 응답부를 나타내는 블록도이다.
- [45] 도 15는 일 실시 예에 따른 전자 장치 및 외부의 서버가 서로 연동함으로써 데이터를 학습하고 판단하는 예시를 나타내는 도면이다.
- [46] 도 16은 본 개시의 일 실시 예에 따른 전자 장치의 제어 방법을 설명하기 위한 흐름도이다.

발명의 실시를 위한 최선의 형태

[47] -

발명의 실시를 위한 형태

- [48] 이하, 본 개시의 다양한 실시 예가 첨부된 도면을 참조하여 기재된다. 그러나, 이는 본 개시에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 실시 예의 다양한 변경(modifications), 균등물(equivalents), 및/또는 대체물(alternatives)을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.
- [49] 본 개시에서, "가진다," "가질 수 있다," "포함한다," 또는 "포함할 수 있다" 등의 표현은 해당 특징(예: 수치, 기능, 동작, 또는 부품 등의 구성요소)의 존재를 가리키며, 추가적인 특징의 존재를 배제하지 않는다.
- [50] 본 개시에서, "A 또는 B," "A 또는/및 B 중 적어도 하나," 또는 "A 또는/및 B 중 하나 또는 그 이상" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. 예를 들면, "A 또는 B," "A 및 B 중 적어도 하나," 또는 "A 또는 B 중 적어도 하나"는, (1) 적어도 하나의 A를 포함, (2) 적어도 하나의 B를 포함, 또는 (3) 적어도 하나의 A 및 적어도 하나의 B 모두를 포함하는 경우를 모두 지칭할 수 있다.
- [51] 본 개시에서 사용된 "제1," "제2," "첫째," 또는 "둘째," 등의 표현들은 다양한

- 구성요소들을, 순서 및/또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다.
- [52] 어떤 구성요소(예: 제1 구성요소)가 다른 구성요소(예: 제2 구성요소)에 "(기능적으로 또는 통신적으로) 연결되어((operatively or communicatively) coupled with/to)" 있다거나 "접속되어(connected to)" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제3 구성요소)를 통하여 연결될 수 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소(예: 제1 구성요소)가 다른 구성요소(예: 제2 구성요소)에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소와 상기 다른 구성요소 사이에 다른 구성요소(예: 제 3 구성요소)가 존재하지 않는 것으로 이해될 수 있다.
- [53] 본 개시에서 사용된 표현 "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, "~에 적합한(suitable for)," "~하는 능력을 가지는(having the capacity to)," "~하도록 설계된(designed to)," "~하도록 변경된(adapted to)," "~하도록 만들어진(made to)," 또는 "~를 할 수 있는(capable of)"과 바꾸어 사용될 수 있다. 용어 "~하도록 구성된(또는 설정된)"은 하드웨어적으로 "특별히 설계된(specifically designed to)" 것만을 반드시 의미하지 않을 수 있다. 대신, 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성된(또는 설정된) 부프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(generic-purpose processor)(예: CPU 또는 application processor)를 의미할 수 있다.
- [54] 본 개시의 다양한 실시 예들에 따른 전자 장치는, 예를 들면, 스마트폰, 태블릿 PC, 이동 전화기, 영상 전화기, 전자책 리더기, 데스크탑 PC, 랩탑 PC, 넷북 컴퓨터, 워크스테이션, 서버, PDA, PMP(portable multimedia player), MP3 플레이어, 의료기기, 카메라, 또는 웨어러블 장치 중 적어도 하나를 포함할 수 있다. 웨어러블 장치는 액세서리형(예: 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용형 장치(head-mounted-device(HMD))), 직물 또는 의류 일체형(예: 전자 의복), 신체 부착형(예: 스킨 패드 또는 문신), 또는 생체 이식형 회로 중 적어도 하나를 포함할 수 있다. 어떤 실시 예들에서, 전자 장치는, 예를 들면, 텔레비전, DVD(digital video disk) 플레이어, 오디오, 냉장고, 에어컨, 청소기, 오븐, 전자레인지, 세탁기, 공기 청정기, 셋톱 박스, 홈 오토메이션 컨트롤 패널, 보안 컨트롤 패널, 미디어 박스(예: 삼성 HomeSync™, 애플TV™, 또는 구글 TV™), 게임 콘솔(예: Xbox™, PlayStation™), 전자 사전, 전자 키, 캠코더, 또는 전자 액자 중 적어도 하나를 포함할 수 있다.
- [55] 다른 실시 예에서, 전자 장치는, 각종 의료기기(예: 각종 휴대용

의료측정기기(혈당 측정기, 심박 측정기, 혈압 측정기, 또는 체온 측정기 등), MRA(magnetic resonance angiography), MRI(magnetic resonance imaging), CT(computed tomography), 촬영기, 또는 초음파기 등), 네비게이션 장치, 위성 항법 시스템(GNSS(global navigation satellite system)), EDR(event data recorder), FDR(flight data recorder), 자동차 인포테인먼트 장치, 선박용 전자 장비(예: 선박용 항법 장치, 자이로 콤팩스 등), 항공 전자기기(avionics), 보안 기기, 차량용 헤드 유닛(head unit), 산업용 또는 가정용 로봇, 드론(drone), 금융 기관의 ATM, 상점의 POS(point of sales), 또는 사물 인터넷 장치 (예: 전구, 각종 센서, 스프링클러 장치, 화재 경보기, 온도조절기, 가로등, 토스터, 운동기구, 온수탱크, 히터, 보일러 등) 중 적어도 하나를 포함할 수 있다.

- [56] 본 개시에서, 사용자라는 용어는 전자 장치를 사용하는 사람 또는 전자 장치를 사용하는 장치(예: 인공지능 전자 장치)를 지칭할 수 있다.
- [57] 도 1a 및 도 1b는 본 개시의 일 실시 예에 따른 복수의 사용자 음성을 처리하는 방법을 설명하기 위한 도면들이다.
- [58] 먼저, 도 1a에 도시된 바와 같이, 전자 장치는 사용자 A가 발화한 사용자 음성을 수신할 수 있다. 예를 들어, 전자 장치는 사용자 A가 발화한 "강남에서 동기 모임 캘린더에 입력해줘."라는 사용자 음성을 수신할 수 있다.
- [59] 전자 장치는 사용자 A의 사용자 음성을 외부 서버로 전송할 수 있다. 이때, 외부 서버는 사용자 음성을 텍스트로 변환하기 위한 제1 서버 및 변환된 텍스트를 바탕으로 사용자 음성에 대한 응답 정보를 제공하기 위한 제2 서버를 포함할 수 있으나, 이는 일 실시 예에 불과할 뿐, 단일 서버로 구현될 수 있다.
- [60] 외부 서버는 STT(Speech to Text) 기술을 통해 전자 장치로부터 수신된 사용자 음성을 텍스트로 변환할 수 있다. 그리고, 외부 서버는 인공지능 알고리즘을 통해 학습된 인공지능 모델에 변환된 텍스트를 입력하여 사용자 음성에 대한 응답 정보를 획득할 수 있다. 이때, 응답 정보는 사용자 음성에 대한 응답을 포함하는 응답 사운드로서, 오디오 데이터로 전자 장치로 전송될 수 있다. 또는, 응답 정보는 제2 서버에 의해 획득된 텍스트 정보로서, TTS(Text to speech) 기술을 이용하여 제1 서버에 의해 응답 사운드로 변환되어 전자 장치(100)로 전송될 수 있다. 다만, 이에 한정되는 것은 아니며, 응답 정보는 음성에 대한 정보 뿐만 아니라 영상에 대한 정보를 더 포함할 수도 있다. 이 경우, 영상에 대한 정보는 사용자 음성에 대한 응답을 포함하는 텍스트, 이미지, URL 등과 같은 정보일 수 있다. 또한, 음성에 대한 정보는 영상에 대한 정보와 동일할 수 있으나, 이는 일 실시 예에 불과할 뿐, 영상에 대한 정보 중 일부만을 포함하거나 영상에 대한 정보를 요약한 정보일 수도 있다.
- [61] 외부 서버는 사용자 음성에 대한 응답 사운드를 전자 장치로 전송할 수 있다. 예를 들어, 외부 서버는 사용자 A의 사용자 음성에 대한 응답으로 "몇일/몇시인가요?"와 같은 응답 사운드를 전자 장치로 전송할 수 있다. 전자 장치는 수신된 응답 사운드를 제공할 수 있다.

- [62] 다만, 외부 서버에서 응답 사운드를 제공하는 기능을 전자 장치가 수행할 수도 있다. 예를 들어, 전자 장치는 사용자 A가 발화한 사용자 음성을 텍스트로 변환하고, 인공지능 모델에 변환된 텍스트를 입력하여 사용자 음성에 대한 응답 정보를 획득하며, 응답 정보를 응답 사운드로 변경하여 사용자 A에게 제공할 수도 있다.
- [63] 이후, 전자 장치는 사용자 B가 발화한 사용자 음성을 수신할 수 있다. 예를 들어, 전자 장치는 사용자 B가 발화한 "오늘 날씨 어때?"라는 사용자 음성을 수신할 수 있다.
- [64] 전자 장치는 사용자 A의 사용자 음성 및 사용자 B의 사용자 음성을 비교하여 사용자의 동일 여부를 판단할 수 있다. 예를 들어, 전자 장치는 사용자 A의 사용자 음성 및 사용자 B의 사용자 음성 각각의 보이스 프린트에 기초하여 사용자의 동일 여부를 판단할 수 있다. 전자 장치는 사용자가 달라진 경우, 사용자 B의 사용자 음성을 스토리지에 저장하고 외부 서버로 전송하지 않을 수 있다.
- [65] 또는, 외부 서버가 사용자의 동일 여부를 판단할 수도 있다. 예를 들어, 전자 장치는 사용자의 동일 여부에 대한 판단 없이, 사용자 B의 사용자 음성을 외부 서버로 전송할 수 있다. 외부 서버는 수신된 사용자 B의 사용자 음성을 이전의 사용자 A의 사용자 음성과 비교하여 사용자의 동일 여부를 판단할 수 있다. 외부 서버는 사용자가 달라진 경우, 사용자 B의 사용자 음성을 스토리지에 저장할 수 있다. 이 경우, 외부 서버는 사용자 B의 사용자 음성에 대한 응답 사운드를 생성하지 않을 수 있다. 전자 장치는 외부 서버로부터 아무런 정보를 수신하지 못하며, 사용자 B의 사용자 음성에 대한 응답 사운드를 제공하지 않을 수 없다. 또는, 외부 서버는 사용자 A에 대한 응답을 제공하는 상황이며, 사용자 B의 대기를 안내하는 응답 사운드를 전자 장치로 전송할 수도 있다. 전자 장치는 수신된 응답 사운드를 사용자 B에게 제공할 수도 있다.
- [66] 이후, 전자 장치는 사용자 A가 발화한 사용자 음성을 수신할 수 있다. 예를 들어, 전자 장치는 사용자 A가 발화한 "이번달 17일 3시."라는 사용자 음성을 수신할 수 있다.
- [67] 전자 장치는 이상과 같은 방법으로 사용자 A의 사용자 음성에 대한 응답 사운드를 제공할 수 있다. 예를 들어, 전자 장치는 사용자 A의 사용자 음성에 대해 "12월 17일 3시로 등록했습니다."와 같은 응답 사운드를 제공할 수 있다.
- [68] 이후, 도 1b에 도시된 바와 같이, 전자 장치는 사용자 A로부터 기설정된 시간 동안 사용자 음성이 입력되지 않으면, 사용자 B가 발화한 "오늘 날씨 어때?"라는 사용자 음성을 외부 서버로 전송할 수 있다. 외부 서버는 이상과 같은 방법으로 사용자 B의 사용자 음성에 대한 "아까 물어보신 오늘 날씨는 맑음입니다."와 같은 응답 사운드를 전자 장치로 전송할 수 있다. 전자 장치는 수신된 응답 사운드를 사용자 B에게 제공할 수 있다. 여기서, 전자 장치는 사용자 B의 존재를 식별하고, 응답 사운드를 제공할 수 있다. 전자 장치는 사용자 B의 존재가

식별되지 않으면, 응답 사운드를 스토리지에 저장하고, 이후 사용자 B의 존재가 식별되면 응답 사운드를 제공할 수도 있다.

- [69] 또는, 외부 서버가 사용자 B의 사용자 음성을 저장하고 있고, 외부 서버가 "12월 17일 3시로 등록했습니다."와 같은 응답 사운드를 전자 장치로 제공하는 경우, 외부 서버는 사용자 A와의 인터랙션이 종료된 것으로 판단할 수 있다. 외부 서버는 사용자 A와의 인터랙션이 종료된 것으로 판단되면, 스토리지에 저장된 사용자 B의 사용자 음성에 대한 "아까 물어보신 오늘 날씨는 맑음입니다."와 같은 응답 사운드를 전자 장치로 전송할 수도 있다.
- [70] 한편, 상술한 바와 같은 외부 서버가 제공하는 학습된 모델은 인공지능 알고리즘 기반으로 학습된 판단 모델로서, 예로, 신경망(Neural Network)을 기반으로 하는 모델일 수 있다. 학습된 판단 모델은 인간의 뇌 구조를 컴퓨터 상에서 모의하도록 설계될 수 있으며 인간의 신경망의 뉴런(neuron)을 모의하는, 가중치를 가지는 복수의 네트워크 노드들을 포함할 수 있다. 복수의 네트워크 노드들은 뉴런이 시냅스(synapse)를 통하여 신호를 주고 받는 뉴런의 시냅틱(synaptic) 활동을 모의하도록 각각 연결 관계를 형성할 수 있다. 또한 학습된 판단 모델은, 일 예로, 신경망 모델, 또는 신경망 모델에서 발전한 딥 러닝 모델을 포함할 수 있다. 딥 러닝 모델에서 복수의 네트워크 노드들은 서로 다른 깊이(또는, 레이어)에 위치하면서 컨볼루션(convolution) 연결 관계에 따라 데이터를 주고 받을 수 있다. 학습된 판단 모델의 예에는 DNN(Deep Neural Network), RNN(Recurrent Neural Network), BRDNN(Bidirectional Recurrent Deep Neural Network) 등이 있을 수 있으나 이에 한정되지 않는다.
- [71] 또한, 전자 장치는 상술한 바와 같이 사용자 음성에 대한 응답을 음성 및 영상으로 제공받기 위하여, 인공지능 전용 프로그램(또는 인공지능 에이전트, Artificial intelligence agent)인 개인 비서 프로그램을 이용할 수 있다. 이때, 개인 비서 프로그램은 AI(Artificial Intelligence) 기반의 서비스를 제공하기 위한 전용 프로그램으로서, 기존의 범용 프로세서(예를 들어, CPU) 또는 별도의 AI 전용 프로세서(예를 들어, GPU 등)에 의해 실행될 수 있다.
- [72] 구체적으로, 기설정된 사용자 입력(예를 들어, 전자 장치를 지칭하는 텍스트에 대응되는 사용자 발화 등)이 입력되거나 전자 장치에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지는 경우, 인공지능 에이전트가 동작(또는 실행)할 수 있다. 그리고, 인공지능 에이전트는 입력된 사용자 음성을 외부 서버로 전송하고, 외부 서버로부터 수신된 음성에 대한 정보를 바탕으로 사용자 음성에 대한 응답 사운드를 스피커를 통해 제공하고, 외부 서버로부터 수신된 영상에 대한 정보를 바탕으로 사용자 음성에 대한 응답을 포함하는 가이드 화면을 제공할 수 있다.
- [73] 물론, 화면 상에 기설정된 사용자 입력이 감지되거나 전자 장치에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지면, 인공지능 에이전트가 동작할 수도 있다. 또는, 인공지능 에이전트는 기설정된

사용자 입력이 감지되거나 전자 장치에 구비된 버튼이 선택되기 이전에 기 실행된 상태일 수도 있다. 이 경우, 기설정된 사용자 입력이 감지되거나 전자 장치에 구비된 버튼이 선택된 이후에는 전자 장치의 인공지능 에이전트가 사용자 음성에 대한 응답을 포함하는 음성 및 영상을 제공할 수 있다. 또한, 인공지능 에이전트는 기설정된 사용자 입력이 감지되거나 전자 장치에 구비된 버튼이 선택되기 이전에 대기 상태일 수 있다. 여기서, 대기 상태란, 인공지능 에이전트의 동작 시작을 제어하기 위해 미리 정의된 사용자 입력이 수신되는 것을 감지하는 상태이다. 인공지능 에이전트가 대기 상태인 동안 기설정된 사용자 입력이 감지되거나 전자 장치에 구비된 버튼이 선택되면, 전자 장치는 인공지능 에이전트를 동작시키고, 사용자 음성에 대한 응답을 포함하는 음성 및 영상을 제공할 수 있다.

- [74] 본 개시의 또 다른 실시 예로, 전자 장치가 직접 인공지능 모델을 이용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득하는 경우, 인공지능 에이전트는 인공지능 모델을 제어하여 음성 및 영상에 대한 정보를 획득할 수 있다. 이때, 인공지능 에이전트는 상술한 외부 서버의 동작을 수행할 수 있다.
- [75]
- [76] 도 2는 본 개시의 일 실시 예에 따른 사용자 음성에 대한 응답 사운드를 제공하는 전자 시스템을 도시한 도면이다. 도 2에 도시된 바와 같이, 전자 시스템은 전자 장치(100) 및 외부 서버(200)를 포함할 수 있다.
- [77] 전자 장치(100)는 인공지능 전용 프로그램을 저장할 수 있다. 여기서, 인공지능 전용 프로그램은 다양한 방식으로 구현될 수 있다.
- [78] 먼저, 제1 방식에 따른 인공지능 전용 프로그램은 사용자 음성을 수신하고, 사용자 음성에 대한 응답 사운드를 제공하기 위해 전자 장치(100)를 이용할 수 있다. 이 경우, 전자 장치(100)는 입력된 사용자 음성을 외부 서버(200)로 전송하고, 외부 서버(200)로부터 사용자 음성에 대한 응답 사운드를 수신하고, 수신된 응답 사운드를 사용자에게 제공할 수 있다.
- [79] 외부 서버(200)는 전자 장치(100)로부터 수신된 사용자 음성으로부터 사용자를 식별할 수 있다. 예를 들어, 외부 서버(200)는 사용자 음성의 보이스 프린트로부터 사용자를 식별할 수 있다. 다만, 이에 한정되는 것은 아니며, 사용자 음성으로 사용자를 식별할 수 있는 방법이라면 어떠한 것이라고 무방하다.
- [80] 그리고, 외부 서버(200)는 사용자 음성의 도메인을 결정할 수 있다. 예를 들어, 외부 서버(200)는 "오늘 날씨 어때?"와 같은 사용자 음성이 수신되면, 수신된 사용자 음성의 도메인이 정보 요청 도메인인 것으로 식별할 수 있다. 또는, 외부 서버(200)는 "최신 음악 들려줘"와 같은 사용자 음성이 수신되면, 수신된 사용자 음성의 도메인이 기기 제어 도메인인 것으로 식별할 수 있다.
- [81] 또한, 외부 서버(200)는 수신된 사용자 음성이 발화 히스토리 정보와

컨텍스트가 일치하는지 판단할 수 있다. 예를 들어, 외부 서버(200)는 "오늘 날씨 어때?"와 같은 사용자 음성이 수신된 후, "내일은?"과 같은 사용자 음성이 수신되면 컨텍스트가 일치하는 것으로 판단할 수 있다. 또는, 외부 서버(200)는 "오늘 날씨 어때?"와 같은 사용자 음성이 수신된 후, "지금 TV에서 뭐해?"와 같은 사용자 음성이 수신되면 컨텍스트가 불일치하는 것으로 판단할 수 있다.

[82] 외부 서버(200)는 식별된 사용자, 수신된 음성의 도메인 및 컨텍스트 일치 여부 등에 기초하여 응답 정보를 생성할 수 있다. 즉, 식별된 사용자, 수신된 음성의 도메인 및 컨텍스트 일치 여부 등의 정보는 사용자 음성과 함께 외부 서버(200)에 저장된 인공지능 모델에 입력될 수 있다.

[83] 다만, 이는 일 실시 예에 불과하고, 인공지능 모델에는 사용자 음성만이 입력될 수도 있다. 예를 들어, 인공지능 모델은 입력된 사용자 음성으로부터 사용자를 식별하는 제1 인공지능 서브 모델 및 사용자 음성으로부터 응답 정보를 생성하는 제2 인공지능 서브 모델을 포함할 수 있다.

[84] 이 경우, 제2 인공지능 서브 모델은 사용자 음성의 도메인 및 컨텍스트 일치 여부 등을 판단하지 않을 수 있다. 다만, 제2 인공지능 서브 모델의 학습 과정에서 이용된 샘플 대화 패턴 등은 일반적으로 도메인 및 컨텍스트가 일치되는 대화 패턴일 수 있다. 그에 따라, 도메인 또는 컨텍스트가 발화 히스토리 정보와 다른 사용자 음성이 입력되면 제2 인공지능 서브 모델은 입력된 사용자 음성을 별도 처리하거나 사용자 음성의 처리를 유예할 수 있다.

[85] 즉, 외부 서버(200)는 인공지능 모델을 이용하는 경우, 사용자 음성의 도메인 및 컨텍스트를 구체적으로 결정하지 않을 수 있다. 다만, 외부 서버(200)는 학습 과정에서 이용된 샘플 대화 패턴 등의 영향으로 사용자 음성의 도메인 및 컨텍스트 등이 실질적으로 반영된 응답 정보를 생성할 수 있다.

[86] 외부 서버(200)는 생성된 응답 정보를 응답 사운드로 변경하고, 응답 사운드를 전자 장치(100)로 전송할 수 있다.

[87] 한편, 제2 방식에 따른 인공지능 전용 프로그램은 사용자 음성을 수신하고, 사용자를 식별하며, 사용자 음성에 대한 응답 사운드를 제공하기 위해 전자 장치(100)를 이용할 수 있다. 이 경우, 전자 장치(100)는 입력된 사용자 음성을 발화한 사용자를 식별하고, 입력된 사용자 음성 및 식별된 사용자 정보를 외부 서버(200)로 전송할 수 있다. 그리고, 전자 장치(100)는 외부 서버(200)로부터 사용자 음성에 대한 응답 사운드를 수신하고, 수신된 응답 사운드를 사용자에게 제공할 수 있다.

[88] 외부 서버(200)는 사용자 음성 및 사용자 정보를 인공지능 알고리즘에 입력하여 응답 정보를 생성할 수 있다. 이때, 외부 서버(200)는 사용자 정보에 기초하여 발화 히스토리 정보를 관리할 수 있다. 예를 들어, 외부 서버(200)는 사용자 A 및 사용자 A의 사용자 음성이 수신되면, 사용자 A의 발화 히스토리 정보 및 사용자 A의 사용자 음성을 인공지능 알고리즘에 입력하여 응답 정보를 생성하고, 사용자 B 및 사용자 B의 사용자 음성이 수신되면, 사용자 B의 발화

히스토리 정보 및 사용자 B의 사용자 음성을 인공지능 알고리즘에 입력하여 응답 정보를 생성할 수 있다.

- [89] 외부 서버(200)는 생성된 응답 정보를 응답 사운드로 변경하고, 응답 사운드를 전자 장치(100)로 전송할 수 있다.
- [90] 한편, 제3 방식에 따른 인공지능 전용 프로그램은 인공지능 모델을 포함하는 프로그램일 수 있다. 이 경우, 전자 장치(100)는 입력된 사용자 음성을 발화한 사용자를 식별하고, 입력된 사용자 음성 및 식별된 사용자 정보를 인공지능 모델에 입력하여 응답 정보를 생성할 수 있다. 전자 장치(100)는 응답 정보를 응답 사운드로 변경하고, 응답 사운드를 스피커를 통해 출력할 수 있다. 즉, 전자 장치(100)는 외부 서버(200) 없이 사용자 음성에 대한 응답 사운드를 제공할 수 있다.
- [91]
- [92] 도 3a는 본 개시의 일 실시 예에 따른 전자 장치(100)의 구성을 설명하기 위한 블록도이다.
- [93] 도 3a에 도시된 바와 같이, 전자 장치(100)는 마이크(110), 스피커(120) 및 프로세서(130)를 포함할 수 있다.
- [94] 전자 장치(100)는 사용자 음성을 수신하고, 수신된 사용자 음성에 따라 다양한 인터랙션을 제공하는 장치일 수 있다. 예를 들어, 전자 장치(100)는 스피커, 노트북 PC, 태블릿 PC, 스마트폰 등일 수 있다. 다만, 이에 한정되는 것은 아니며, 전자 장치(100)는 사용자 음성을 수신하고, 그에 따른 응답 사운드를 제공할 수 있는 장치이면 어떠한 것이라도 무방하다.
- [95] 또한, 전자 장치(100)는 스피커(120) 대신 디스플레이를 포함하는 장치일 수도 있다. 이 경우, 전자 장치(100)는 사용자 음성을 수신하고, 수신된 사용자 음성에 따른 인터랙션을 디스플레이를 통해 제공할 수도 있다. 그 밖에 전자 장치(100)는 사용자 음성을 수신하고, 수신된 사용자 음성에 대한 응답 정보를 제공할 수 있는 장치라면 어떠한 장치라도 무방하다.
- [96] 마이크(110)는 사용자 음성이나 기타 소리를 입력받아 오디오 데이터로 변환하기 위한 구성이다. 마이크(110)는 프로세서(130)와 전기적으로 연결되며, 프로세서(130)의 제어에 의해 사용자 음성을 수신할 수 있다.
- [97] 마이크(110)는 복수의 서브 마이크를 포함할 수 있다. 예를 들어, 마이크(110)는 전자 장치(100)의 전후좌우에 각각 하나의 서브 마이크를 포함할 수 있다. 다만, 이에 한정되는 것은 아니며, 전자 장치(100)는 하나의 마이크(110)만을 포함할 수도 있다.
- [98] 스피커(120)는 각종 오디오 데이터 뿐만 아니라 각종 알림 음이나 음성 메시지 등을 출력하는 구성요소이다. 스피커(120)는 프로세서(130)와 전기적으로 연결되며, 프로세서(130)의 제어에 의해 응답 사운드를 출력할 수 있다.
- [99] 스피커(120)는 복수의 서브 스피커를 포함할 수 있다. 예를 들어, 스피커(120)는 전자 장치(100)의 좌우에 각각 하나의 서브 스피커를 포함할 수 있다. 다만, 이에

한정되는 것은 아니며, 전자 장치(100)는 하나의 스피커(120)만을 포함할 수도 있다.

[100] 프로세서(130)는 전자 장치(100)의 동작을 전반적으로 제어한다.

[101] 일 실시 예에 따라 프로세서(130)는 디지털 시그널 프로세서(digital signal processor(DSP), 마이크로 프로세서(microprocessor), TCON(Time controller)으로 구현될 수 있다. 다만, 이에 한정되는 것은 아니며, 중앙처리장치(central processing unit(CPU)), MCU(Micro Controller Unit), MPU(micro processing unit), 컨트롤러(controller), 어플리케이션 프로세서(application processor(AP)), 또는 커뮤니케이션 프로세서(communication processor(CP)), ARM 프로세서 중 하나 또는 그 이상을 포함하거나, 해당 용어로 정의될 수 있다. 또한, 프로세서(130)는 프로세싱 알고리즘이 내장된 SoC(System on Chip), LSI(large scale integration)로 구현될 수도 있고, FPGA(Field Programmable gate array) 형태로 구현될 수도 있다.

[102] 프로세서(130)는 마이크(110)를 통해 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 사용자를 식별하고 인공지능 알고리즘을 통해 학습된 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 스피커(120)를 통해 제공할 수 있다.

[103] 여기서, 인공지능 모델은 전자 장치(100)에 저장되어 있을 수 있고, 이 경우 프로세서(130)는 제1 사용자 음성을 인공지능 모델에 입력하여 제1 응답 사운드를 획득할 수 있다.

[104] 또는, 인공지능 모델은 외부 서버(200)에 저장되어 있을 수 있고, 이 경우 프로세서(130)는 제1 사용자 음성을 외부 서버(200)로 전송하고, 제1 사용자 음성을 인공지능 모델에 입력하여 획득된 제1 응답 사운드를 외부 서버로(200)로부터 수신할 수 있다.

[105] 프로세서(130)는 마이크(110)를 통해 제2 사용자 음성이 입력되면, 제2 사용자 음성을 발화한 사용자를 식별하고, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 스피커(120)를 통해 제공할 수 있다.

[106] 예를 들어, 프로세서(130)는 제1 사용자 음성을 발화한 사용자가 발화한 제2 사용자 음성을 수신할 수 있고, 이 경우 프로세서(130)는 인공지능 모델에 제2 사용자 음성 뿐만 아니라 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공할 수 있다. 즉, 제2 응답 사운드는 현재 입력된 사용자 음성 뿐만 아니라 그 이전에 입력된 사용자 음성 및 사용자 음성에 따른 응답 정보를 모두 반영하여 획득될 수 있다.

[107] 여기서, 전자 장치(100)는 카메라를 더 포함하고, 프로세서(130)는 카메라에 의해 촬영된 영상 및 마이크를 통해 입력된 사용자 음성 중 적어도 하나에 기초하여 사용자를 식별할 수 있다.

[108] 구체적으로, 프로세서(130)는 제1 사용자 음성이 입력되는 시간 구간 동안

영상에 포함된 사용자의 얼굴을 식별하여 제1 사용자 음성을 발화한 사용자를 식별할 수 있다.

[109] 또는, 프로세서(130)는 제1 사용자 음성이 입력되는 시간 구간 동안 영상에 포함된 사용자의 입모양에 기초하여 제1 사용자 음성을 발화한 사용자를 식별할 수도 있다.

[110] 예를 들어, 프로세서(130)는 카메라를 통해 실시간 촬영을 수행하며, 제1 사용자 음성이 입력되는 시간 구간 동안 실시간 촬영 중인 영상에 포함된 복수의 사용자 중 입모양의 변경이 있는 사용자를 제1 사용자 음성을 발화한 사용자로서 식별할 수 있다.

[111] 그리고, 프로세서(130)는 제2 사용자 음성이 입력되는 시간 구간 동안 영상에 포함된 식별된 사용자의 입모양이 변경되면, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일한 것으로 식별할 수 있다.

[112] 또는, 프로세서(130)는 스피커(120)를 이용해 사용자의 동일 여부를 식별할 수도 있다. 예를 들어, 스피커(120)는 복수의 서브 스피커를 포함하며, 프로세서(130)는 제1 사용자 음성이 입력되면 복수의 서브 스피커를 이용하여 제1 사용자 음성을 발화한 사용자의 위치를 식별할 수 있다.

[113] 그리고, 프로세서(130)는 제2 사용자 음성이 입력되면 복수의 서브 스피커를 이용하여 제2 사용자 음성을 발화한 사용자의 위치를 식별할 수 있으며, 이전에 식별된 사용자의 위치와 비교하여 사용자의 변경 여부를 판단할 수 있다.

[114] 또는, 프로세서(130)는 사용자 음성의 보이스 프린트를 통해 사용자를 식별할 수도 있다. 예를 들어, 전자 장치(100)는 프로세서(130)와 전기적으로 연결된 스토리지를 더 포함하고, 프로세서(130)는 복수의 사용자 음성 각각의 주파수 분석을 수행하고, 복수의 사용자 각각의 주파수 특성을 스토리지에 저장할 수 있다. 이후, 프로세서(130)는 제1 사용자 음성이 입력되면 제1 사용자 음성의 주파수 분석을 수행하고, 스토리지에 저장된 정보에 기초하여 제1 사용자 음성을 발화한 사용자를 식별할 수 있다. 프로세서(130)는 제2 사용자 음성에 대하여도 동일한 동작을 수행하고, 식별된 사용자의 동일 여부를 판단할 수 있다.

[115] 한편, 보이스 프린트를 이용하는 동작은 외부 서버(200)에서 수행될 수도 있다. 이 경우, 전자 장치(100)는 사용자를 식별하는 동작을 수행하지 않고, 입력된 사용자 음성을 외부 서버(200)로 전송할 수 있다. 외부 서버(200)는 수신된 사용자 음성의 주파수 분석을 수행하여 수신된 사용자 음성을 발화한 사용자를 식별할 수 있다.

[116] 또는, 프로세서(130)는 입력된 사용자 음성을 발화한 사용자를 식별하지 않고, 남/여 구분, 아이/성인/노인 구분 및 출신 지역 등을 식별할 수 있다. 예를 들어, 프로세서(130)는 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 사용자가 남자 성인이고, 표준어를 사용하는 점을 식별할 수 있다. 그리고, 프로세서(130)는 제2 사용자 음성이 입력되면, 제2 사용자 음성을 발화한

사용자가 여자 성인이고, 표준어를 사용하는 점을 식별할 수 있다. 이 경우, 프로세서(130)는 사용자가 동일하지 않음을 식별할 수 있다.

[117] 한편, 사용자 음성에 대한 남/여 구분, 아이/성인/노인 구분 및 출신 지역 등을 식별은 외부 서버(200)에서 수행될 수도 있다. 이 경우, 전자 장치(100)는 사용자를 식별하는 동작을 수행하지 않고, 입력된 사용자 음성을 외부 서버(200)로 전송할 수 있다. 외부 서버(200)는 수신된 사용자 음성을 이용하여 남/여 구분, 아이/성인/노인 구분 및 출신 지역 등을 식별하고, 사용자의 동일 여부를 판단할 수도 있다.

[118] 한편, 이상에서 프로세서(100)는 사용자 음성이 입력된 시점으로부터 기설정된 시간 동안은 입력된 사용자 음성 및 사용자 음성에 대한 응답 사운드를 발화 히스토리 정보로서 스토리지에 저장할 수 있다. 예를 들어, 프로세서(100)는 "내일 날씨 어때?"와 같은 사용자 음성이 입력되면, 사용자 음성에 대한 "내일 날씨는 맑음입니다."와 같은 응답 사운드를 제공하고, 사용자 음성이 입력된 시점으로부터 10분 동안 인터랙션 정보를 발화 히스토리 정보로서 스토리지에 저장할 수 있다. 프로세서(130)는 사용자 음성이 입력된 시점으로부터 10분 내에 동일한 사용자로부터 추가 사용자 음성이 입력되면, 인공지능 모델에 추가 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 응답 사운드를 제공할 수 있다. 또는, 프로세서(130)는 사용자 음성이 입력된 시점으로부터 10분이 경과한 후 동일한 사용자로부터 추가 사용자 음성이 입력되면, 추가 사용자 음성을 발화한 사용자와 최초 사용자 음성을 발화한 사용자의 동일 여부를 판단하지 않고, 인공지능 모델에 추가 사용자 음성만을 입력하여 획득된 응답 사운드를 제공할 수 있다. 즉, 프로세서(130)는 사용자 음성이 입력된 시점으로부터 기설정된 시간 동안 추가 사용자 음성이 입력된 경우에만 추가 사용자 음성을 발화한 사용자와 최초 사용자 음성을 발화한 사용자의 동일 여부를 판단할 수 있다.

[119] 프로세서(130)는 사용자 음성이 입력된 시점으로부터 기설정된 시간 이내에 추가 사용자 음성이 입력되면, 기설정된 시간의 시작 시점을 추가 사용자 음성이 입력된 시점으로 변경할 수 있다. 예를 들어, 프로세서(130)는 오후 6시에 제1 사용자 음성이 입력되고, 제1 사용자 음성이 입력된 오후 6시부터 기설정된 시간 10분 내인 오후 6시 5분에 제2 사용자 음성이 입력되면, 오후 6시 5분부터 기설정된 시간 10분 이후인 오후 6시 15분까지 발화 히스토리 정보를 저장할 수 있다. 여기서, 발화 히스토리 정보는 제1 사용자 음성, 제1 사용자 음성에 대한 제1 응답 사운드, 제2 사용자 음성, 제2 사용자 음성에 대한 제2 응답 사운드를 포함할 수 있다. 즉, 프로세서(130)는 사용자가 연속적으로 사용자 음성을 발화하여 발화 히스토리 정보를 저장하는 시간이 연장되면, 연장된 시간까지의 발화 히스토리 정보가 스토리지에 저장된 상태로 유지될 수 있다.

[120] 다만, 이에 한정되는 것은 아니며, 사용자가 연속적으로 사용자 음성을 발화하여 발화 히스토리 정보를 저장하는 시간이 연장되더라도, 연장된

시간까지의 발화 히스토리 정보가 스토리지에 저장된 상태로 유지되지 않을 수도 있다. 예를 들어, 프로세서(130)는 사용자가 연속적으로 사용자 음성을 발화하여 발화 히스토리 정보를 저장하는 시간이 연장되는 경우, 현재 시점으로부터 기설정된 시간 이전의 시점부터 현재 시점까지의 사용자 음성 및 응답 사운드만을 발화 히스토리 정보로서 저장할 수 있다. 즉, 프로세서(130)는 발화 히스토리 정보 중 발화 시점으로부터 기설정된 시간이 경과한 사용자 음성 및 사운드 제공 시점으로부터 기설정된 시간이 경과한 응답 사운드는 스토리지에서 삭제할 수도 있다.

- [121] 한편, 전자 장치(100)가 인공지능 모델을 저장하고 있는 경우, 스토리지는 사용자 별 발화 히스토리 정보를 시간 순으로 구분하여 저장할 수 있다. 예를 들어, 스토리지는 사용자 A의 2017년 1월 1일의 발화 히스토리 정보, 2017년 1월 5일의 발화 히스토리 정보 및 사용자 B의 2017년 1월 4일의 발화 히스토리 정보 등을 저장할 수 있다. 프로세서(130)는 스토리지에 저장된 발화 히스토리 정보 중 기설정된 시간이 경과한 발화 히스토리 정보를 외부 서버(200)로 전송할 수 있다. 프로세서(130)는 필요에 따라 외부 서버(200)로 전송한 발화 히스토리 정보를 외부 서버(200)로부터 수신할 수 있다. 프로세서(130)는 외부 서버(200)로 발화 히스토리 정보를 전송하기 전, 발화 히스토리 정보의 키워드 등을 추출하고, 추출된 키워드를 스토리지에 저장한 후 발화 히스토리 정보를 외부 서버(200)로 전송할 수도 있다. 추출된 키워드는 추후 사용자가 과거 발화 히스토리 정보를 검색하는 경우에 이용될 수 있다.
- [122] 예를 들어, 사용자가 "5년 전쯤 제주도에서 가려고 검색했던 음식점이 뭐였더라?"와 같은 사용자 음성을 발화하면, 프로세서(130)는 제주도 및 음식점을 키워드로 5년 이전의 키워드를 검색하고, 대응되는 발화 히스토리 정보를 외부 서버(200)로부터 수신할 수 있다. 프로세서(130)는 외부 서버(200)로부터 수신한 발화 히스토리 정보로부터 사용자가 원하는 정보를 제공할 수 있다.
- [123] 한편, 전자 장치(100)는 스토리지를 더 포함하고, 프로세서(130)는 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 제2 사용자 음성을 스토리지에 저장하고, 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자 음성이 입력되지 않으면, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 스피커(120)를 통해 제공할 수 있다.
- [124] 예를 들어, 프로세서(130)는 사용자 A가 제1 사용자 음성을 발화한 시점으로부터 기설정된 시간 이내에 사용자 B가 제2 사용자 음성을 발화하면 사용자의 동일 여부를 판단하고, 사용자가 동일하지 않다고 판단되면 사용자 B의 제2 사용자 음성을 스토리지에 저장할 수 있다. 즉, 프로세서(130)는 사용자 B의 제2 사용자 음성에 대한 응답 사운드를 제공을 연기할 수 있다. 그리고, 프로세서(130)는 사용자 A가 제1 사용자 음성을 발화한 시점으로부터 기설정된

시간이 경과되기 전까지 사용자 A의 추가 사용자 음성이 입력되지 않으면, 제1 사용자 음성을 발화한 시점으로부터 기설정된 시간이 경과된 후 제2 사용자 음성에 대한 응답 사운드를 제공할 수 있다. 여기서, 제2 사용자 음성에 대한 응답 사운드는 사용자 A에 의한 발화 히스토리 정보가 삭제된 상태에서 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 응답 사운드일 수 있다.

[125] 한편, 프로세서(130)는 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 제1 사용자를 식별하고 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 스피커(120)를 통해 제공하며, 제2 사용자 음성이 입력되면, 제2 사용자 음성을 발화한 제2 사용자를 식별하고, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 스피커(120)를 통해 제1 응답 사운드와 구분되게 제공할 수 있다. 즉, 프로세서(130)는 복수의 사용자 음성을 구분하여 실시간으로 처리하며, 처리된 결과의 제공 상태를 서로 다르게 할 수 있다.

[126] 예를 들어, 프로세서(130)는 사용자 A의 "오늘 날씨 어때?"에 대한 제1 응답 사운드로 "오늘 날씨는 맑음입니다."를 남자 목소리로 제공하고, 사용자 B의 "오늘 밤에 영화 뭐해?"에 대한 제2 응답 사운드로 "오늘 밤에는 영화를 하지 않습니다."를 여자 목소리로 제공할 수 있다. 여기서, 제2 응답 사운드는 사용자 A의 발화 히스토리 정보를 반영하지 않고, 사용자 B의 사용자 음성에만 기초하여 획득된 응답 사운드일 수 있다.

[127] 다만, 이에 한정되는 것은 아니며, 프로세서(130)는 얼마든지 다른 방법으로 사용자 음성에 대한 응답 정보의 제공 상태를 다르게 제공할 수 있다. 예를 들어, 전자 장치(100)는 프로세서(130)와 전기적으로 연결된 디스플레이를 더 포함하고, 프로세서(130)는 사용자 A의 "오늘 날씨 어때?"에 대한 제1 응답으로 "오늘 날씨는 맑음입니다."를 디스플레이를 통해 디스플레이하고, 사용자 B의 "오늘 밤에 영화 뭐해?"에 대한 제2 응답 사운드로 "오늘 밤에는 영화를 하지 않습니다."를 스피커(120)를 통해 제공할 수도 있다.

[128] 또는, 프로세서(130)는 사용자 A의 "오늘 날씨 어때?"에 대한 제1 응답으로 "오늘 날씨는 맑음입니다."와 사용자 B의 "오늘 밤에 영화 뭐해?"에 대한 제2 응답으로 "오늘 밤에는 영화를 하지 않습니다."를 모두 디스플레이할 수 있으며, 사용자들이 응답의 대상을 식별할 수 있도록 제1 응답과 제2 응답의 배경 화면 색상을 다르게 디스플레이할 수도 있다.

[129] 한편, 프로세서(130)는 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 인공지능 모델에 제1 추가 사용자 음성 및 제1 사용자 음성에 대응되는 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를 스피커(120)를 통해 제공하고, 제2 사용자에 의해 제2 추가 사용자 음성이 입력되면, 인공지능 모델에 제2 추가 사용자 음성 및 제2 사용자 음성에 대응되는 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 스피커(120)를 통해 제1 추가 응답 사운드와 구분되게 제공할 수 있다.

- [130] 그리고, 프로세서(130)는 제1 추가 응답 사운드를 제공하며 제1 사용자에게 대응되는 제1 UI를 디스플레이를 통해 디스플레이하고, 제2 추가 응답 사운드를 제공하며 제2 사용자에게 대응되는 제2 UI를 디스플레이를 통해 디스플레이할 수 있다.
- [131] 예를 들어, 프로세서(130)는 제1 추가 응답 사운드를 제공하며 제1 사용자의 이미지를 디스플레이를 통해 디스플레이하고, 제2 추가 응답 사운드를 제공하며 제2 사용자의 이미지를 디스플레이를 통해 디스플레이할 수 있다.
- [132] 한편, 인공지능 알고리즘이 저장된 외부 서버(200)를 이용하여 사용자별로 응답 사운드를 제공하는 경우, 프로세서(130)는 사용자 음성을 식별된 사용자 정보와 함께 외부 서버(200)로 전송할 수 있다. 외부 서버(200)는 수신된 사용자 정보에 기초하여 사용자 별로 발화 히스토리 정보를 생성하고, 인공지능 모델에 사용자 음성과 대응되는 발화 히스토리 정보를 입력하여 획득된 응답 사운드를 전자 장치(100)로 전송할 수 있다. 이때, 외부 서버(200)는 응답 사운드 뿐만 아니라 응답 사운드의 대상에 대한 사용자 정보를 전자 장치(100)로 전송할 수 있다.
- [133] 전자 장치(100)는 수신된 사용자 정보에 기초하여 응답 사운드를 다른 사용자의 응답 사운드와 구분되게 제공할 수 있다.
- [134] 한편, 프로세서(130)는 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 제1 사용자 음성에 대응되는 제1 도메인 및 제2 사용자 음성에 대응되는 제2 도메인을 획득하고, 제1 도메인 및 제2 도메인이 동일하면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 스피커(120)를 통해 제공할 수도 있다.
- [135] 예를 들어, 프로세서(130)는 사용자 A의 "오늘 날씨 어때?"와 같은 제1 사용자 음성에 대한 제1 응답 사운드로 "오늘 오후에는 비가 옵니다."를 스피커(120)를 통해 제공하고, 사용자 B의 "우산 챙겨야 해?"와 같은 제2 사용자 음성이 입력되면, 사용자의 동일 여부를 판단할 수 있다. 프로세서(130)는 사용자가 동일하지 않다고 판단되면, 제1 사용자 음성의 도메인 및 제2 사용자 음성의 도메인이 모두 정보 요청 도메인임을 획득할 수 있다. 프로세서(130)는 제1 사용자 음성 및 제2 사용자 음성의 도메인이 동일하다고 판단되면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 "네. 오늘은 우산을 챙기세요."와 같은 제2 응답 사운드를 제공할 수 있다.
- [136] 상술한 예에서 사용자 B가 "음악 틀어줘."와 같은 제2 사용자 음성을 발화하면, 프로세서(130)는 사용자의 동일 여부를 판단할 수 있다. 프로세서(130)는 사용자가 동일하지 않다고 판단되면, 제1 사용자 음성의 도메인이 정보 요청 도메인임을 획득하고, 제2 사용자 음성의 도메인이 기기 제어 도메인임을 획득할 수 있다. 프로세서(130)는 제1 사용자 음성 및 제2 사용자 음성의 도메인이 다르다고 판단되면, 제2 사용자 음성의 처리를 연기할 수 있다. 또는, 프로세서(130)는 인공지능 모델에 제2 사용자 음성만을 입력하여 획득된 "네.

최신곡을 틀겠습니다."와 같은 제2 응답 사운드를 제1 응답 사운드와 구분되게 제공할 수도 있다.

[137] 또한, 프로세서(130)는 도메인 뿐만 아니라 컨텍스트를 더 고려할 수도 있다. 예를 들어, 프로세서(130)는 제1 사용자 음성과 제2 사용자 음성이 도메인이 동일하더라도 컨텍스트가 다르면, 제2 사용자 음성의 처리를 연기할 수 있다. 또는, 프로세서(130)는 인공지능 모델에 제2 사용자 음성만을 입력하여 획득된 제2 응답 사운드를 제1 응답 사운드와 구분되게 제공할 수도 있다.

[138] 한편, 프로세서(130)는 제2 사용자 음성이 제1 사용자 음성이 입력된 시점으로부터 기설정된 제1 시간 내에 입력되고, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하며, 제1 사용자 음성에 대응되는 제1 도메인 및 제2 사용자 음성에 대응되는 제2 도메인이 서로 다르면, 제1 응답 사운드의 제공을 생략하고 제1 사용자 음성을 스토리지에 저장하며, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 스피커를 통해 제공할 수 있다. 여기서, 기설정된 제1 시간은 상술한 발화 히스토리 정보의 저장이 유지되는 시간보다 짧을 수 있다.

[139] 예를 들어, 프로세서(130)는 사용자 A의 "TV 틀어줘."와 같은 제1 사용자 음성이 입력된 후 1초 내에 동일 사용자 A의 제1 사용자 음성과 도메인이 다른 "참. 오늘 약속 장소가 어디지?"와 같은 제2 사용자 음성이 입력되면, 제1 사용자 음성을 스토리지에 저장하고, 제2 사용자 음성에 대한 제3 응답 사운드로 "시청 앞입니다."를 제공할 수 있다.

[140] 그리고, 프로세서(130)는 사용자로부터 제2 도메인에 대응되는 사용자 음성이 기설정된 제2 시간 이상 입력되지 않으면, 인공지능 모델에 스토리지에 저장된 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 스피커(120)를 통해 제공할 수 있다. 여기서, 기설정된 제2 시간은 상술한 발화 히스토리 정보의 저장이 유지되는 시간과 동일하고, 기설정된 제1 시간보다 긴 시간일 수 있다. 다만, 이에 한정되는 것은 아니며, 기설정된 제2 시간은 상술한 발화 히스토리 정보의 저장이 유지되는 시간과는 상관이 없고, 기설정된 제1 시간보다 긴 시간일 수도 있다.

[141] 또는, 상술한 예에서, 프로세서(130)는 제2 사용자 음성이 제1 사용자 음성에 의한 컨텍스트와 일치하는지를 판단하고, 컨텍스트가 다르면 제1 응답 사운드의 제공을 생략하고 제1 사용자 음성을 스토리지에 저장하며, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 스피커를 통해 제공할 수도 있다.

[142] 한편, 전자 장치(100) 또는 외부 서버(200)에 저장된 인공지능 알고리즘은 사용자 음성의 도메인을 획득하기 위한 제1 인공지능 알고리즘 및 사용자 음성 간 컨텍스트의 일치 여부를 판단하기 위한 제2 인공지능 알고리즘 중 적어도 하나를 포함할 수도 있다. 이 경우, 응답 정보의 획득 전에 도메인 및 컨텍스트의 일치 여부에 대한 판단이 선행될 수 있다.

- [143] 또는, 전자 장치(100) 또는 외부 서버(200)에 저장된 인공지능 알고리즘은 사용자 음성의 도메인을 획득하거나 사용자 음성 간 컨텍스트의 일치 여부를 판단하기 위한 별도의 인공지능 알고리즘을 포함하지 않을 수도 있다. 이 경우, 인공지능 알고리즘은 그 자체로서 도메인 및 컨텍스트 중 적어도 하나를 반영한 응답 사운드를 제공할 수도 있다.
- [144] 도 3b는 본 개시의 일 실시 예에 따른 전자 장치(100)의 세부 구성을 나타내는 블록도이다. 전자 장치(100)는 마이크(110), 스피커(120) 및 프로세서(130)를 포함할 수 있다. 또한, 도 3b에 따르면, 전자 장치(100)는 통신부(140), 스토리지(145), 디스플레이(150), 사용자 인터페이스부(155), 카메라(160), 오디오 처리부(170), 비디오 처리부(180), 버튼(190)을 더 포함할 수도 있다. 도 3b에 도시된 구성요소들 중 도 3a에 도시된 구성요소와 중복되는 부분에 대해서는 자세한 설명을 생략하도록 한다.
- [145] 프로세서(130)는 스토리지(145)에 저장된 각종 프로그램을 이용하여 전자 장치(100)의 동작을 전반적으로 제어한다.
- [146] 구체적으로, 프로세서(130)는 RAM(131), ROM(132), 메인 CPU(133), 그래픽 처리부(134), 제1 내지 n 인터페이스(135-1 ~ 135-n), 버스(136)를 포함한다.
- [147] RAM(131), ROM(132), 메인 CPU(133), 그래픽 처리부(134), 제1 내지 n 인터페이스(135-1 ~ 135-n) 등은 버스(136)를 통해 서로 연결될 수 있다.
- [148] 제1 내지 n 인터페이스(135-1 내지 135-n)는 상술한 각종 구성요소들과 연결된다. 인터페이스들 중 하나는 네트워크를 통해 외부 장치와 연결되는 네트워크 인터페이스가 될 수도 있다.
- [149] 메인 CPU(133)는 스토리지(145)에 액세스하여, 스토리지(145)에 저장된 O/S를 이용하여 부팅을 수행한다. 그리고, 스토리지(145)에 저장된 각종 프로그램 등을 이용하여 다양한 동작을 수행한다.
- [150] ROM(132)에는 시스템 부팅을 위한 명령어 세트 등이 저장된다. 턴온 명령이 입력되어 전원이 공급되면, 메인 CPU(133)는 ROM(132)에 저장된 명령어에 따라 스토리지(145)에 저장된 O/S를 RAM(131)에 복사하고, O/S를 실행시켜 시스템을 부팅시킨다. 부팅이 완료되면, 메인 CPU(133)는 스토리지(145)에 저장된 각종 어플리케이션 프로그램을 RAM(131)에 복사하고, RAM(131)에 복사된 어플리케이션 프로그램을 실행시켜 각종 동작을 수행한다.
- [151] 그래픽 처리부(134)는 연산부(미도시) 및 렌더링부(미도시)를 이용하여 아이콘, 이미지, 텍스트 등과 같은 다양한 객체를 포함하는 화면을 생성한다. 연산부(미도시)는 수신된 제어 명령에 기초하여 화면의 레이아웃에 따라 각 객체들이 표시될 좌표값, 형태, 크기, 컬러 등과 같은 속성값을 연산한다. 렌더링부(미도시)는 연산부(미도시)에서 연산한 속성값에 기초하여 객체를 포함하는 다양한 레이아웃의 화면을 생성한다. 렌더링부(미도시)에서 생성된 화면은 디스플레이(150)의 디스플레이 영역 내에 표시된다.
- [152] 한편, 상술한 프로세서(130)의 동작은 스토리지(145)에 저장된 프로그램에

의해 이루어질 수 있다.

- [153] 스토리지(145)는 전자 장치(100)를 구동시키기 위한 O/S(Operating System) 소프트웨어 모듈, 인공지능 모듈, 인공지능 전용 프로그램 모듈, 발화 히스토리 모듈, 사용자 식별 모듈, 도메인 식별 모듈 및 컨텍스트 식별 모듈 등과 같이 다양한 데이터를 저장한다.
- [154] 통신부(140)는 다양한 유형의 통신방식에 따라 다양한 유형의 외부 장치와 통신을 수행하는 구성이다. 통신부(140)는 와이파이 칩(141), 블루투스 칩(142), 무선 통신 칩(143), NFC 칩(144) 등을 포함한다. 프로세서(130)는 통신부(140)를 이용하여 각종 외부 장치와 통신을 수행한다.
- [155] 와이파이 칩(141), 블루투스 칩(142)은 각각 WiFi 방식, 블루투스 방식으로 통신을 수행한다. 와이파이 칩(141)이나 블루투스 칩(142)을 이용하는 경우에는 SSID 및 세션 키 등과 같은 각종 연결 정보를 먼저 송수신하여, 이를 이용하여 통신 연결한 후 각종 정보들을 송수신할 수 있다. 무선 통신 칩(143)은 IEEE, 지그비(zigbee), 3G(3rd Generation), 3GPP(3rd Generation Partnership Project), LTE(Long Term Evolution) 등과 같은 다양한 통신 규격에 따라 통신을 수행하는 칩을 의미한다. NFC 칩(144)은 135kHz, 13.56MHz, 433MHz, 860~960MHz, 2.45GHz 등과 같은 다양한 RF-ID 주파수 대역들 중에서 13.56MHz 대역을 사용하는 NFC(Near Field Communication) 방식으로 동작하는 칩을 의미한다.
- [156] 또한, 통신부(140)는 HDMI, MHL, USB, DP, 썬더볼트, RGB, D-SUB, DVI 등과 같은 유선 통신 인터페이스를 더 포함할 수 있다. 프로세서(130)는 통신부(140)의 유선 통신 인터페이스를 통해 외부 서버(200)와 연결될 수도 있다.
- [157] 디스플레이(150)는 LCD(Liquid Crystal Display), OLED(Organic Light Emitting Diodes) 디스플레이, PDP(Plasma Display Panel) 등과 같은 다양한 형태의 디스플레이로 구현될 수 있다. 디스플레이(150) 내에는 a-si TFT, LTPS(low temperature poly silicon) TFT, OTFT(organic TFT) 등과 같은 형태로 구현될 수 있는 구동 회로, 백라이트 유닛 등도 함께 포함될 수 있다. 한편, 디스플레이(150)는 터치 감지부와 결합하여 터치 스크린으로 구현될 수 있다.
- [158] 사용자 인터페이스부(155)는 다양한 사용자 인터랙션(interaction)을 수신한다. 여기서, 사용자 인터페이스부(155)는 전자 장치(100)의 구현 예에 따라 다양한 형태로 구현 가능하다. 예를 들어, 사용자 인터페이스부(155)는 전자 장치(100)에 구비된 버튼, 사용자 음성을 수신하는 마이크, 사용자 모션을 감지하는 카메라 등일 수 있다. 또는, 전자 장치(100)가 터치 기반의 단말 장치로 구현되는 경우 사용자 인터페이스부(155)는 터치패드와 상호 레이어 구조를 이루는 터치 스크린 형태로 구현될 수도 있다. 이 경우, 사용자 인터페이스부(155)는 상술한 디스플레이(150)로 사용될 수 있게 된다.
- [159] 오디오 처리부(170)는 오디오 데이터에 대한 처리를 수행하는 구성요소이다. 오디오 처리부(170)에서는 오디오 데이터에 대한 디코딩이나 증폭, 노이즈 필터링 등과 같은 다양한 처리가 수행될 수 있다.

- [160] 비디오 처리부(180)는 비디오 데이터에 대한 처리를 수행하는 구성요소이다. 비디오 처리부(180)에서는 비디오 데이터에 대한 디코딩, 스케일링, 노이즈 필터링, 프레임 레이트 변환, 해상도 변환 등과 같은 다양한 이미지 처리를 수행할 수 있다.
- [161] 버튼(190)은 전자 장치(100)의 본체 외관의 전면부나 측면부, 배면부 등의 임의의 영역에 형성된 기계적 버튼, 터치 패드, 휠 등과 같은 다양한 유형의 버튼이 될 수 있다.
- [162] 이상과 같은 방법을 통해 프로세서(130)는 사용자가 변경되더라도 컨텍스트가 유지된 응답 정보를 제공할 수 있다.
- [163] 이하에서는 도면을 통해 전자 장치(100)의 동작을 구체적으로 설명한다.
- [164]
- [165] 도 4a 및 도 4b는 본 개시의 일 실시 예에 따른 복수의 사용자 중 한 명의 사용자 음성 처리의 연기를 설명하기 위한 도면들이다.
- [166] 도 4a에 도시된 바와 같이, 프로세서(130)는 사용자 A의 "강남에서 동기모임 캘린더에 입력해줘."와 같은 제1 사용자 음성이 입력되면, 사용자 A를 식별하고 제1 사용자 음성에 대해 "몇일/몇시 인가요?"와 같은 제1 응답 사운드를 제공할 수 있다.
- [167] 그리고, 프로세서(130)는 사용자 B의 "불고기 피자 구매해줘."와 같은 제2 사용자 음성이 입력되면, 사용자 B를 식별하고, 식별된 사용자 B가 사용자 A와 다른 사람이므로 제2 사용자 음성을 스토리지(145)에 저장할 수 있다. 이때, 프로세서(130)는 제2 사용자 음성 뿐만 아니라 제2 사용자 음성을 발화한 사용자 B에 대한 정보도 스토리지(145)에 저장할 수도 있다.
- [168] 그리고, 프로세서(130)는 사용자 A의 "7월 7일이야."와 같은 제3 사용자 음성이 입력되면, 사용자 A를 식별하고, 제1 사용자 음성을 발화한 사용자 A와 동일하므로 제3 사용자 음성에 대해 "7월 7일로 등록했습니다."와 같은 제2 응답 사운드를 제공할 수 있다.
- [169] 이후, 도 4b에 도시된 바와 같이, 프로세서(130)는 사용자 A의 인터랙션이 종료된 것으로 판단되면, 제2 사용자 음성에 대해 "죄송합니다. 어제는 아버지와 대화 태스크 중이라 대응을 못해드렸네요. 구매 요청하신 피자는 구매리스트에 넣었습니다. 몇 개 주문할까요?"와 같은 제3 응답 사운드를 제공할 수 있다. 이때, 프로세서(130)는 제2 사용자가 주변에 있음을 감지한 후, 제3 응답 사운드를 제공할 수 있다.
- [170] 여기서, 프로세서(130)는 사용자 A가 감지되지 않거나 사용자 A로부터 기설정된 시간 동안 사용자 음성이 입력되지 않는 경우에 사용자 A의 인터랙션이 종료된 것으로 판단할 수 있다.
- [171] 이후, 프로세서(130)는 사용자 B의 "상황 이해해. 괜찮아. 3개 주문해줘."와 같은 제4 사용자 음성이 입력되면, 사용자 B를 식별하고, 제2 사용자 음성을 발화한 사용자 B와 동일하므로 제4 사용자 음성에 대해 "네 처리하였습니다."와

같은 제4 응답 사운드를 제공할 수 있다.

[172]

[173] 도 5a 내지 도 5c는 본 개시의 일 실시 예에 따른 복수의 사용자 음성의 실시간 동시 처리 방법을 설명하기 위한 도면들이다.

[174] 도 5a에 도시된 바와 같이, 프로세서(130)는 사용자 A의 "결혼식 언제야?"와 같은 제1 사용자 음성이 입력되면, 사용자 A를 식별하고 제1 사용자 음성에 대해 "7월 7일입니다."와 같은 제1 응답 사운드를 제공할 수 있다.

[175] 그리고, 프로세서(130)는 사용자 B의 "오늘 날씨 어때?"와 같은 제2 사용자 음성이 입력되면, 사용자 B를 식별하고, 식별된 사용자 B가 사용자 A와 다른 사람이므로 제1 사용자 음성 및 제1 응답 사운드를 포함하는 제1 발화 히스토리 정보와 무관하게 제2 사용자 음성만을 고려하여 "오늘 날씨는 흐림입니다."와 같은 제2 응답 사운드를 제공할 수 있다.

[176] 이때, 프로세서(130)는 제1 응답 사운드와는 다른 방식으로 제2 응답 사운드를 제공할 수 있다. 예를 들어, 프로세서(130)는 제1 응답 사운드를 남자 목소리로 제공하고, 제2 응답 사운드를 여자 목소리로 제공할 수 있다.

[177] 또는, 프로세서(130)는 식별된 사용자가 스토리지(145)에 기저장되어 있는 사용자라면, 사용자 이름, ID 등을 포함시켜 응답 사운드를 제공할 수도 있다. 예를 들어, 사용자 A가 "Park"이라는 ID로 스토리지(145)에 기저장되어 있고, 사용자 B가 "Kim"이라는 ID로 스토리지(145)에 기저장되어 있으면, 프로세서(130)는 "Park님. 7월 7일입니다."와 같은 제1 응답 사운드 및 "Kim님. 오늘 날씨는 흐림입니다."와 같은 제2 응답 사운드를 제공할 수도 있다.

[178] 또는, 프로세서(130)는 도 5b 및 도 5c와 같이, 제공되는 응답 사운드의 대상에 따라 서로 다른 색상을 디스플레이하도록 디스플레이(150)를 제어할 수도 있다. 또는, 전자 장치(130)는 LED를 포함하고, 프로세서(130)는 제공되는 응답 사운드의 대상에 따라 서로 다른 색상의 광을 방출하도록 LED를 제어할 수도 있다. 또는, 전자 장치(130)는 복수의 LED를 포함하고, 프로세서(130)는 제공되는 응답 사운드의 대상에 따라 복수의 LED 중 하나가 턴온되도록 LED를 제어할 수도 있다.

[179] 그리고, 프로세서(130)는 사용자 A의 "7월 14일로 변경해줘."와 같은 제3 사용자 음성이 입력되면, 사용자 A를 식별하고, 제1 사용자 음성을 발화한 사용자 A와 동일하므로 제1 발화 히스토리 정보에 기초하여 제3 사용자 음성에 대해 "변경했습니다."와 같은 제3 응답 사운드를 제공할 수 있다.

[180] 한편, 이상에서 도 4a 및 도 4b는 복수의 사용자 음성의 입력이 있더라도 한 명의 사용자 음성에 대하여만 처리하는 실시 예이고, 도 5a 내지 도 5c는 복수의 사용자 음성의 입력을 각각 구분하여 처리하는 실시 예이다. 전자 장치(100)는 도 4a 및 도 4b와 같은 개인 모드 및 도 5a 내지 도 5c와 같은 공동 모드를 제공하며, 프로세서(130)는 사용자의 모드 선택에 따라 개인 모드 또는 공동 모드로 동작할 수 있다.

[181]

[182] 도 6은 본 개시의 일 실시 예에 따른 복수의 사용자와의 통합 인터랙션을 설명하기 위한 도면이다.

[183] 도 6에 도시된 바와 같이, 프로세서(130)는 사용자 A의 "강남에서 동기모임 캘린더에 입력해줘."와 같은 제1 사용자 음성이 입력되면, 사용자 A를 식별하고 제1 사용자 음성에 대해 "몇일/몇시 인가요?"와 같은 제1 응답 사운드를 제공할 수 있다.

[184] 그리고, 프로세서(130)는 사용자 B의 "7월 7일 월요일이 동기모임이야."와 같은 제2 사용자 음성이 입력되면, 사용자 B를 식별하고, 식별된 사용자 B가 사용자 A와 다른 사람이나, 제1 사용자 음성 및 제2 사용자 음성의 컨텍스트가 일치하므로 제1 사용자 음성 및 제1 응답 사운드를 포함하는 제1 발화 히스토리 정보 및 제2 사용자 음성을 고려하여 "7월 7일로 등록했습니다."와 같은 제2 응답 사운드를 제공할 수 있다.

[185] 프로세서(130)는 제1 사용자 음성 및 제2 사용자 음성의 컨텍스트가 일치하는지 불명확하면, 사용자들에게 문의하는 응답 사운드를 제공할 수도 있다. 예를 들어, 프로세서(130)는 제1 사용자 음성 및 제2 사용자 음성의 컨텍스트가 일치하는지 불명확하면, "7월 7일 월요일 동기모임이 강남에서 동기모임인가요?"와 같은 제3 응답 사운드를 제공할 수도 있다.

[186] 한편, 프로세서(130)는 도 6과 같이 통합 인터랙션을 수행하다가 복수의 사용자의 사용자 음성 간 컨텍스트가 변경되면, 공동 모드로 동작할 수도 있다. 또는, 프로세서(130)는 공동 모드로 동작하다가 복수의 사용자의 사용자 음성 간 컨텍스트가 일치하면, 통합 인터랙션을 수행할 수도 있다.

[187]

[188] 도 7은 본 개시의 일 실시 예에 따른 동일 사용자의 복수의 도메인 각각에 대응하는 복수의 사용자 음성이 입력된 경우의 동작을 설명하기 위한 도면이다.

[189] 도 7에 도시된 바와 같이, 프로세서(130)는 사용자 A의 "큰방 TV 시청할 때면 다른 방 불 꺼주고 이 방은 시원하고 어둡게 해줘."와 같은 제1 사용자 음성이 입력되고, 곧바로 "참. 다음 주말에 결혼식 일정 등록해줘."와 같은 제2 사용자 음성이 입력되면, 제1 사용자 음성 및 제2 사용자 음성을 발화한 사용자 A를 식별하고 늦게 입력된 제2 사용자 음성에 대해 "무슨 요일, 몇시인가요?"와 같은 제1 응답 사운드를 제공할 수 있다. 이때, 프로세서(130)는 제1 사용자 음성을 스토리지(145)에 저장할 수 있다. 또는, 프로세서(130)는 제1 사용자 음성 뿐만 아니라 제1 사용자 음성을 발화한 사용자 A에 대한 정보도 스토리지(145)에 저장할 수도 있다.

[190] 그리고, 프로세서(130)는 사용자 A의 "토요일 12시."와 같은 제3 사용자 음성이 입력되면, 사용자 A를 식별하고, 제1 사용자 음성을 발화한 사용자 A와 동일하므로 제2 사용자 음성 및 제1 응답 사운드를 포함하는 제1 발화 히스토리 정보에 기초하여 제3 사용자 음성에 대해 "네. 결혼식 일정 등록했습니다."와

같은 제2 응답 사운드를 제공할 수 있다.

- [191] 프로세서(130)는 사용자 음성에 따른 특정 동작을 완료함 따라 사용자와의 인터랙션이 종료된 것으로 판단할 수 있다. 이상의 예에서 프로세서(130)는 결혼식 일정 등록을 완료함에 따라 사용자 A의 제1 인터랙션이 종료된 것으로 판단할 수 있다. 다만, 이에 한정되는 것은 아니며, 프로세서(130)는 사용자가 감지되지 않거나, 동일 사용자로부터 기설정된 시간 이상 사용자 음성이 입력되지 않은 경우에도 사용자와의 인터랙션이 종료된 것으로 판단할 수도 있다. 또는, 프로세서(130)는 사용자의 종료 명령에 따라 사용자와의 인터랙션이 종료된 것으로 판단할 수도 있다.
- [192] 프로세서(130)는 사용자와의 인터랙션이 종료된 것으로 판단되면, 스토리지(145)에 저장된 사용자 음성을 획득하고, 새로운 인터랙션을 시작할 수 있다. 이상의 예에서 프로세서(130)는 스토리지(145)에 저장된 제1 사용자 음성을 획득하고, 제1 사용자 음성에 대해 "이전에 요청하셨던 내용에서 거실과 복도 불도 함께 끌까요?"와 같은 제3 응답 사운드를 제공할 수 있다.
- [193] 이때, 프로세서(130)는 스토리지(145)에 저장된 사용자 음성을 발화한 사용자가 감지되는 경우, 새로운 인터랙션을 시작할 수도 있다. 이상의 예에서 프로세서(130)는 스토리지(145)에 저장된 제1 사용자 음성을 획득하더라도 제1 사용자 음성을 발화한 사용자 A가 감지되지 않으면, 새로운 인터랙션을 시작하지 않을 수 있다.
- [194] 이후, 프로세서(130)는 사용자 A의 "거실은 끄고 복도는 켜줘."와 같은 제4 사용자 음성이 입력되면, 제4 사용자 음성에 대해 "네. 알겠습니다."와 같은 제4 응답 사운드를 제공할 수 있다.
- [195]
- [196] 도 8a 및 도 8b는 본 개시의 일 실시 예에 따른 사용자 음성을 발화한 사용자를 식별하기 위한 방법을 설명하기 위한 도면들이다.
- [197] 도 8a에 도시된 바와 같이, 프로세서(130)는 사용자 음성이 입력되면, 카메라(160)에 의해 촬영된 영상(810)으로부터 사용자를 식별할 수 있다. 또는, 프로세서(130)는 사용자 음성이 입력되면, 카메라(160)에 의해 촬영된 영상(810)으로부터 사용자의 입모양이 움직이는 사용자를 사용자 음성을 발화한 사용자로서 식별할 수도 있다.
- [198] 또는, 도 8b에 도시된 바와 같이, 스피커(110)는 서로 이격되어 설치된 복수의 서브 스피커(110-1, 110-2, 110-3, 110-4)를 포함하고, 프로세서(130)는 복수의 서브 스피커(110-1, 110-2, 110-3, 110-4)를 이용하여 사용자 음성이 발화된 제1 위치를 추정할 수 있다. 그리고, 프로세서(130)는 추가 사용자 음성이 입력되면 추가 사용자 음성이 발화된 제2 위치를 추정하고, 제2 위치가 제1 위치로부터 기설정된 거리 이내이면 동일한 사용자로 식별할 수 있다.
- [199]
- [200] 도 9a 및 도 9b는 본 개시의 일 실시 예에 따른 인공지능 모델의 추가 학습을

설명하기 위한 도면들이다.

- [201] 상술한 바와 같이 전자 장치(100) 또는 외부 서버(200)에 저장된 인공지능 모델은 샘플 대화 패턴 등을 이용하여 기 학습된 상태일 수 있다. 이후, 인공지능 모델은 사용자의 인터랙션에 기초하여 재학습될 수 있다. 여기서, 재학습은 샘플 대화 패턴에 의해 학습된 인공지능 모델은 사용자와의 인터랙션에 기초하여 추가로 학습하여 인공지능 모델을 업데이트하는 것을 의미한다.
- [202] 도 9a에 도시된 바와 같이, 프로세서(130)는 사용자 A의 "지금 TV에서 뭐해?"와 같은 제1 사용자 음성이 입력되면, 제1 사용자 음성에 대해 "뉴스9 합니다."와 같은 제1 응답 사운드를 제공할 수 있다.
- [203] 그리고, 프로세서(130)는 사용자 A의 "영화는 안해?"와 같은 제2 사용자 음성이 입력되면, 제1 사용자 음성, 제1 응답 사운드 및 제2 사용자 음성에 기초하여 인공지능 모델의 재학습을 수행할 수 있다.
- [204] 예를 들어, 재학습된 인공지능 모델은 사용자가 영화를 선호하는 점이 반영된 인공지능 모델일 수 있다.
- [205] 도 9b는 도 9a와 동일한 상황을 전제하였으며, 도 9a는 재학습 전의 인공지능 모델을 이용한 경우의 도면이고 도 9b는 재학습된 인공지능 모델을 이용한 경우의 도면이다.
- [206] 도 9b에 도시된 바와 같이, 재학습된 인공지능 모델을 이용하면, 프로세서(130)는 사용자 A의 "지금 TV에서 뭐해?"와 같은 제1 사용자 음성이 입력되면, 제1 사용자 음성에 대해 "영화는 안하고 미드합니다."와 같은 제2 응답 사운드를 제공할 수 있다.
- [207] 한편, 도 9a 및 도 9b에서는 설명의 편의를 위해 전자 장치(100)가 인공지능 모델을 저장한 것으로 설명하였으나, 인공지능 모델이 외부 서버(200)에 저장된 경우에도 이상의 재학습 방법은 동일하게 적용될 수 있다.
- [208]
- [209] 도 10은 본 개시의 일 실시 예에 따른 인공지능 모델이 저장된 전자 장치(100)의 동작을 설명하기 위한 흐름도이다.
- [210] 먼저, 프로세서(130)는 사용자 음성을 수신한다(S1010). 그리고, 프로세서(130)는 사용자를 감지한다(S1020). 그리고, 프로세서(130)는 이전 사용자와 동일한 사용자인지 판단하여(S1030), 동일한 사용자이면 발화 히스토리 정보에 기초하여 사용자 음성을 처리한다(S1050).
- [211] 프로세서(130)는 동일한 사용자가 아니면 이전 도메인과 동일한 도메인인지 판단하고(S1040), 동일한 도메인이면 발화 히스토리 정보에 기초하여 사용자 음성을 처리한다(S1050). 프로세서(130)는 동일한 도메인이 아니면 발화 히스토리 정보를 고려하지 않고 사용자 음성을 처리한다(S1060).
- [212] 그리고, 프로세서(130)는 음성 처리에 따른 응답 사운드를 제공한다(S1070).
- [213]
- [214] 도 11은 본 개시의 일 실시 예에 따른 전자 장치(100) 및 인공지능 모델이

저장된 외부 서버(200)의 동작을 설명하기 위한 시퀀스도이다.

- [215] 먼저, 전자 장치(100)는 사용자 음성을 수신하고(S1110), 사용자를 감지한다(S1120). 그리고, 전자 장치(100)는 사용자 정보 및 사용자 음성을 외부 서버(200)로 전송한다(S1130).
- [216] 외부 서버(200)는 사용자 정보에 기초하여 사용자 음성에 대응되는 응답 정보를 생성한다(S1140). 그리고, 외부 서버(200)는 응답 정보를 전자 장치(100)로 전송한다(S1150).
- [217] 그리고, 전자 장치(100)는 응답 사운드를 제공한다(S1160).
- [218]
- [219] 도 12는 본 개시의 일 실시 예에 따른 사용자 문의에 대한 응답을 제공하는 인공지능 모델을 학습하고 이용하기 위한 외부 서버(1300)의 구성을 나타내는 블록도이다.
- [220] 도 12를 참조하면, 외부 서버(1300)는 학습부(1310) 및 응답부(1320) 중 적어도 하나를 포함할 수 있다. 도 12의 외부 서버(1300)는 도 2의 외부 서버(200)에 대응될 수 있다.
- [221] 학습부(1310)는 학습 데이터를 이용하여 사용자 음성에 대한 응답을 제공하기 위한 기준을 갖는 인공지능 모델을 생성 또는 학습시킬 수 있다. 학습부(1310)는 수집된 학습 데이터를 이용하여 판단 기준을 갖는 판단 모델을 생성할 수 있다.
- [222] 일 예로, 학습부(1310)는 사용자 음성을 학습 데이터로 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득하도록 학습할 수 있다. 또한, 학습부(1310)는 사용자 음성 및 전자 장치의 상태 정보를 학습 데이터로서 이용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 제공하기 위한 인공지능 모델을 생성, 학습 또는 갱신시킬 수 있다.
- [223] 응답부(1320)는 소정의 데이터를 학습된 인공지능 모델의 입력 데이터로 사용하여, 소정의 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득할 수 있다.
- [224] 일 예로, 응답부(1320)는 사용자 음성 및 전자 장치의 상태 정보를 학습된 인공지능 모델의 입력 데이터로 사용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득(또는, 추정, 추론)할 수 있다.
- [225] 본 개시의 일 실시 예로, 학습부(1310) 및 응답부(1320)는 외부 서버(1300)에 포함될 수 있으나, 이는 일 실시 예에 불과할 뿐, 전자 장치(100) 내부에 탑재될 수 있다. 구체적으로, 학습부(1310)의 적어도 일부 및 응답부(1320)의 적어도 일부는, 소프트웨어 모듈로 구현되거나 적어도 하나의 하드웨어 칩 형태로 제작되어 전자 장치(100)에 탑재될 수 있다. 예를 들어, 학습부(1310) 및 응답부(1320) 중 적어도 하나는 인공 지능(AI; artificial intelligence)을 위한 전용 하드웨어 칩 형태로 제작될 수도 있고, 또는 기존의 범용 프로세서(예: CPU 또는 application processor) 또는 그래픽 전용 프로세서(예: GPU)의 일부로 제작되어 전술한 각종 전자 장치에 탑재될 수도 있다. 이때, 인공 지능을 위한 전용

하드웨어 칩은 확률 연산에 특화된 전용 프로세서로서, 기존의 범용 프로세서보다 병렬처리 성능이 높아 기계 학습과 같은 인공지능 분야의 연산 작업을 빠르게 처리할 수 있다. 학습부(1310) 및 응답부(1320)가 소프트웨어 모듈(또는, 인스트럭션(instruction) 포함하는 프로그램 모듈)로 구현되는 경우, 소프트웨어 모듈은 컴퓨터로 읽을 수 있는 판독 가능한 비일시적 판독 가능 기록매체(non-transitory computer readable media)에 저장될 수 있다. 이 경우, 소프트웨어 모듈은 OS(Operating System)에 의해 제공되거나, 소정의 애플리케이션에 의해 제공될 수 있다. 또는, 소프트웨어 모듈 중 일부는 OS(Operating System)에 의해 제공되고, 나머지 일부는 소정의 애플리케이션에 의해 제공될 수 있다.

[226] 이 경우, 학습부(1310) 및 응답부(1320)는 하나의 전자 장치에 탑재될 수도 있으며, 또는 별개의 전자 장치들에 각각 탑재될 수도 있다. 예를 들어, 학습부(1310) 및 응답부(1320) 중 하나는 전자 장치(100)에 포함되고, 나머지 하나는 외부의 서버에 포함될 수 있다. 또한, 학습부(1310) 및 응답부(1320)는 유선 또는 무선으로 통하여, 학습부(1310)가 구축한 모델 정보를 응답부(1320)로 제공할 수도 있고, 학습부(1320)로 입력된 데이터가 추가 학습 데이터로서 학습부(1310)로 제공될 수도 있다.

[227]

[228] 도 13은 본 개시의 일 실시 예에 따른 학습부(1310)를 나타내는 블록도이다.

[229] 도 13을 참조하면, 일부 실시 예에 따른 학습부(1310)는 학습 데이터 획득부(1310-1) 및 모델 학습부(1310-4)를 포함할 수 있다. 또한, 학습부(1310)는 학습 데이터 전처리부(1310-2), 학습 데이터 선택부(1310-3) 및 모델 평가부(1310-5) 중 적어도 하나를 선택적으로 더 포함할 수 있다.

[230] 학습 데이터 획득부(1310-1)는 사용자 음성에 대한 응답을 획득하기 위한 인공지능 모델에 필요한 학습 데이터를 획득할 수 있다. 본 개시의 실시 예로, 학습 데이터 획득부(1310-1)는 사용자 음성을 학습 데이터로서 획득할 수 있다. 또한, 학습 데이터 획득부(1310-1)는 사용자 음성에 대한 구체적인 응답을 획득하기 위해 필요한 정보(예로, 전자 장치의 상태 정보)를 학습 데이터로서 획득할 수 있다. 학습 데이터는 학습부(1310) 또는 학습부(1310)의 제조사가 수집 또는 테스트한 데이터가 될 수도 있다.

[231] 모델 학습부(1310-4)는 학습 데이터를 이용하여, 인공지능 모델이 사용자 음성에 대한 응답을 획득하는 기준을 갖도록 학습시킬 수 있다. 예로, 모델 학습부(1310-4)는 학습 데이터 중 적어도 일부를 사용자 음성에 대한 응답을 획득하기 위한 기준으로 이용하는 지도 학습(supervised learning)을 통하여, 인공지능 모델을 학습시킬 수 있다. 또는, 모델 학습부(1310-4)는, 예를 들어, 별다른 지도 없이 학습 데이터를 이용하여 스스로 학습함으로써, 응답을 제공하기 위한 기준을 발견하는 비지도 학습(unsupervised learning)을 통하여, 인공지능 모델을 학습시킬 수 있다. 또한, 모델 학습부(1310-4)는, 예를 들어,

학습에 따른 응답 제공의 결과가 올바른 지에 대한 피드백을 이용하는 강화 학습(reinforcement learning)을 통하여, 인공지능 모델을 학습시킬 수 있다. 또한, 모델 학습부(1310-4)는, 예를 들어, 오류 역전파법(error back-propagation) 또는 경사 하강법(gradient descent)을 포함하는 학습 알고리즘 등을 이용하여 인공지능 모델을 학습시킬 수 있다

- [232] 또한, 모델 학습부(1310-4)는 입력 데이터를 이용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 제공하기 위하여 어떤 학습 데이터를 이용해야 하는지에 대한 선별 기준을 학습할 수도 있다.
- [233] 모델 학습부(1310-4)는 미리 구축된 인공지능 모델이 복수 개가 존재하는 경우, 입력된 학습 데이터와 기본 학습 데이터의 관련성이 큰 인공지능 모델을 학습할 인공지능 모델로 결정할 수 있다. 이 경우, 기본 학습 데이터는 데이터의 타입별로 기 분류되어 있을 수 있으며, 인공지능 모델은 데이터의 타입별로 미리 구축되어 있을 수 있다. 예를 들어, 기본 학습 데이터는 학습 데이터가 생성된 지역, 학습 데이터가 생성된 시간, 학습 데이터의 크기, 학습 데이터의 장르, 학습 데이터의 생성자, 학습 데이터 내의 오브젝트의 종류 등과 같은 다양한 기준으로 기 분류되어 있을 수 있다.
- [234] 인공지능 모델이 학습되면, 모델 학습부(1310-4)는 학습된 인공지능 모델을 저장할 수 있다. 이 경우, 모델 학습부(1310-4)는 학습된 인공지능 모델을 외부 서버(1300)의 메모리에 저장할 수 있다. 또는, 모델 학습부(1310-4)는 학습된 인공지능 모델을 외부 서버(1300)와 유선 또는 무선 네트워크로 연결되는 서버 또는 전자 장치의 메모리에 저장할 수도 있다.
- [235] 학습부(1310)는 인공지능 모델의 응답 결과를 향상시키거나, 인공지능 모델의 생성에 필요한 자원 또는 시간을 절약하기 위하여, 학습 데이터 전처리부(1310-2) 및 학습 데이터 선택부(1310-3)를 더 포함할 수도 있다.
- [236] 학습 데이터 전처리부(1310-2)는 사용자 음성에 대한 응답을 포함하는 정보를 획득하기 위한 학습에 획득된 데이터가 이용될 수 있도록, 획득된 데이터를 전처리할 수 있다. 학습 데이터 전처리부(1310-2)는 모델 학습부(1310-4)가 사용자 음성에 대한 응답을 포함하는 정보를 획득하기 위하여 획득된 데이터를 이용할 수 있도록, 획득된 데이터를 기 설정된 포맷으로 가공할 수 있다. 예를 들어, 학습 데이터 전처리부(1310-2)는 입력된 사용자 음성 중 인공지능 모델이 응답을 제공하 때 필요없는 텍스트(예를 들어, 부사, 감탄사 등)를 제거할 수 있다.
- [237] 학습 데이터 선택부(1310-3)는 학습 데이터 획득부(1310-1)에서 획득된 데이터 또는 학습 데이터 전처리부(1310-2)에서 전처리된 데이터 중에서 학습에 필요한 데이터를 선택할 수 있다. 선택된 학습 데이터는 모델 학습부(1310-4)에 제공될 수 있다. 학습 데이터 선택부(1310-3)는 기 설정된 선별 기준에 따라, 획득되거나 전처리된 데이터 중에서 학습에 필요한 학습 데이터를 선택할 수 있다. 또한, 학습 데이터 선택부(1310-3)는 모델 학습부(1310-4)에 의한 학습에 의해 기 설정된 선별 기준에 따라 학습 데이터를 선택할 수도 있다.

- [238] 학습부(1310)는 인공지능 모델의 응답 결과를 향상시키기 위하여, 모델 평가부(1310-5)를 더 포함할 수도 있다.
- [239] 모델 평가부(1310-5)는 인공지능 모델에 평가 데이터를 입력하고, 평가 데이터로부터 출력되는 응답 결과가 소정 기준을 만족하지 못하는 경우, 모델 학습부(1310-4)로 하여금 다시 학습하도록 할 수 있다. 이 경우, 평가 데이터는 인공지능 모델을 평가하기 위한 기 정의된 데이터일 수 있다.
- [240] 예를 들어, 모델 평가부(1310-5)는 평가 데이터에 대한 학습된 인공지능 모델의 응답 결과 중에서, 응답 결과가 정확하지 않은 평가 데이터의 개수 또는 비율이 미리 설정된 임계치를 초과하는 경우 소정 기준을 만족하지 못한 것으로 평가할 수 있다.
- [241] 한편, 학습된 인공지능 모델이 복수 개가 존재하는 경우, 모델 평가부(1310-5)는 각각의 학습된 인공지능 모델에 대하여 소정 기준을 만족하는지를 평가하고, 소정 기준을 만족하는 모델을 최종 인공지능 모델로서 결정할 수 있다. 이 경우, 소정 기준을 만족하는 모델이 복수 개인 경우, 모델 평가부(1310-5)는 평가 점수가 높은 순으로 미리 설정된 어느 하나 또는 소정 개수의 모델을 최종 인공지능 모델로서 결정할 수 있다.
- [242]
- [243] 도 14는 본 개시의 일 실시 예에 따른 응답부(1320)를 나타내는 블록도이다.
- [244] 도 14를 참조하면, 일부 실시 예에 따른 응답부(1320)는 입력 데이터 획득부(1320-1) 및 응답 결과 제공부(1320-4)를 포함할 수 있다.
- [245] 또한, 응답부(1320)는 입력 데이터 전처리부(1320-2), 입력 데이터 선택부(1320-3) 및 모델 갱신부(1320-5) 중 적어도 하나를 선택적으로 더 포함할 수 있다.
- [246] 입력 데이터 획득부(1320-1)는 사용자 음성에 대한 응답을 획득하기 위해 필요한 데이터를 획득할 수 있다. 응답 결과 제공부(1320-4)는 입력 데이터 획득부(1320-1)에서 획득된 입력 데이터를 입력 값으로 학습된 인공지능 모델에 적용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득할 수 있다. 응답 결과 제공부(1320-4)는 후술할 입력 데이터 전처리부(1320-2) 또는 입력 데이터 선택부(1320-3)에 의해 선택된 데이터를 입력 값으로 인공지능 모델에 적용하여 응답 결과를 획득할 수 있다. 응답 결과는 인공지능 모델에 의해 결정될 수 있다.
- [247] 일 실시 예로, 응답 결과 제공부(1320-4)는 입력 데이터 획득부(1320-1)에서 획득한 사용자 음성 및 전자 장치를 학습된 인공지능 모델 적용하여 사용자 음성에 대한 응답을 포함하는 음성 및 영상에 대한 정보를 획득할 수 있다.
- [248] 응답부(1320)는 인공지능 모델의 응답 결과를 향상시키거나, 응답 결과의 제공을 위한 자원 또는 시간을 절약하기 위하여, 입력 데이터 전처리부(1320-2) 및 입력 데이터 선택부(1320-3)를 더 포함할 수도 있다.
- [249] 입력 데이터 전처리부(1320-2)는 사용자 음성에 대한 응답을 획득하기 위해

획득된 데이터가 이용될 수 있도록, 획득된 데이터를 전처리할 수 있다. 입력 데이터 전처리부(1320-2)는 응답 결과 제공부(1320-4)가 사용자 음성에 대한 응답을 획득하기 위하여 획득된 데이터를 이용할 수 있도록, 획득된 데이터를 기 정의된 포맷으로 가공할 수 있다.

[250] 입력 데이터 선택부(1320-3)는 입력 데이터 획득부(1320-1)에서 획득된 데이터 또는 입력 데이터 전처리부(1320-2)에서 전처리된 데이터 중에서 응답 제공에 필요한 데이터를 선택할 수 있다. 선택된 데이터는 응답 결과 제공부(1320-4)에게 제공될 수 있다. 입력 데이터 선택부(1320-3)는 응답 제공을 위한 기 설정된 선별 기준에 따라, 획득되거나 전처리된 데이터 중에서 일부 또는 전부를 선택할 수 있다. 또한, 입력 데이터 선택부(1320-3)는 모델 학습부(1310-4)에 의한 학습에 의해 기 설정된 선별 기준에 따라 데이터를 선택할 수도 있다.

[251] 모델 갱신부(1320-5)는 응답 결과 제공부(1320-4)에 의해 제공되는 응답 결과에 대한 평가에 기초하여, 인공지능 모델이 갱신되도록 제어할 수 있다. 예를 들어, 모델 갱신부(1320-5)는 응답 결과 제공부(1320-4)에 의해 제공되는 응답 결과를 모델 학습부(1310-4)에게 제공함으로써, 모델 학습부(1310-4)가 인공지능 모델을 추가 학습 또는 갱신하도록 요청할 수 있다.

[252]

[253] 도 15는 일 실시 예에 따른 전자 장치(100) 및 외부의 서버(S)가 서로 연동함으로써 데이터를 학습하고 판단하는 예시를 나타내는 도면이다.

[254] 도 15를 참조하면, 외부의 서버(S)는 사용자 음성에 대한 응답을 포함하는 정보르 획득하기 위한 기준을 학습할 수 있으며, 전자 장치(100)는 서버(S)에 의한 학습 결과에 기초하여 사용자 음성에 대한 응답을 음성 및 영상을 통해 제공할 수 있다.

[255] 이 경우, 서버(S)의 모델 학습부(1310-4)는 도 13에 도시된 학습부(1310)의 기능을 수행할 수 있다. 즉, 서버(S)의 모델 학습부(1310-4)는 사용자 음성에 대한 응답을 획득하기 위하여 어떤 텍스트 정보 혹은 전자 장치의 상태 정보를 이용할지, 상기 정보를 이용하여 사용자 음성에 대한 응답을 어떻게 획득할지에 관한 기준을 학습할 수 있다.

[256] 또한, 전자 장치(100)의 응답 결과 제공부(1320-4)는 입력 데이터 선택부(1320-3)에 의해 선택된 데이터를 서버(S)에 의해 생성된 인공지능 모델에 적용하여 사용자 음성에 대한 응답을 획득할 수 있다. 또는, 전자 장치(100)의 응답 결과 제공부(1320-4)는 서버(S)에 의해 생성된 인공지능 모델을 서버(S)로부터 수신하고, 수신된 인공지능 모델을 이용하여 사용자 음성에 대한 응답을 획득할 수 있다.

[257]

[258] 도 16은 본 개시의 일 실시 예에 따른 전자 장치의 제어 방법을 설명하기 위한 흐름도이다.

- [259] 먼저, 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 사용자를 식별한다(S1610). 그리고, 인공지능 알고리즘을 통해 학습된 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 제공한다(S1620). 그리고, 제2 사용자 음성이 입력되면, 제2 사용자 음성을 발화한 사용자를 식별한다(S1630). 그리고, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공한다(S1640).
- [260] 여기서, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 제2 사용자 음성을 저장하는 단계 및 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자 음성이 입력되지 않으면, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 제공하는 단계를 더 포함할 수 있다.
- [261] 한편, 제1 사용자 음성을 발화한 사용자를 식별하는 단계(S1610)는 제1 사용자 음성이 입력되면, 제1 사용자 음성을 발화한 제1 사용자를 식별하고, 제1 응답 사운드를 제공하는 단계(S1620)는 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 제공하며, 제2 사용자 음성을 발화한 사용자를 식별하는 단계(S1630)는 제2 사용자 음성을 발화한 제2 사용자를 식별하고, 제어 방법은 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 제1 응답 사운드와 구분되게 제공하는 단계를 더 포함할 수 있다.
- [262] 여기서, 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 인공지능 모델에 제1 추가 사용자 음성 및 제1 사용자 음성에 대응되는 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를 제공하는 단계 및 제2 사용자에 의해 제2 추가 사용자 음성이 입력되면, 인공지능 모델에 제2 추가 사용자 음성 및 제2 사용자 음성에 대응되는 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 제1 추가 응답 사운드와 구분되게 제공하는 단계를 더 포함할 수 있다.
- [263] 그리고, 제1 추가 응답 사운드를 제공하는 단계는 제1 추가 응답 사운드를 제공하며 제1 사용자에 대응되는 제1 UI를 디스플레이하고, 제2 추가 응답 사운드를 제1 추가 응답 사운드와 구분되게 제공하는 단계는 제2 추가 응답 사운드를 제공하며 제2 사용자에 대응되는 제2 UI를 디스플레이할 수 있다.
- [264] 한편, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 제1 사용자 음성에 대응되는 제1 도메인 및 제2 사용자 음성에 대응되는 제2 도메인을 획득하는 단계 및 제1 도메인 및 제2 도메인이 동일하면, 인공지능 모델에 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공하는 단계를 더 포함할 수 있다.
- [265] 또한, 제어 방법은 제2 사용자 음성이 제1 사용자 음성이 입력된 시점으로부터 기설정된 제1 시간 내에 입력되고, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일하며, 제1 사용자 음성에 대응되는 제1

도메인 및 제2 사용자 음성에 대응되는 제2 도메인이 서로 다르면, 제1 응답 사운드의 제공을 생략하고 제1 사용자 음성을 저장하며, 인공지능 모델에 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 제공할 수 있다.

[266] 여기서, 사용자로부터 제2 도메인에 대응되는 사용자 음성이 기설정된 제2 시간 이상 입력되지 않으면, 인공지능 모델에 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 제공하는 단계를 더 포함할 수 있다.

[267] 한편, 사용자를 식별하는 단계(S1610, S1630)는 촬영된 영상 및 입력된 사용자 음성 중 적어도 하나에 기초하여 사용자를 식별할 수 있다.

[268] 여기서, 제1 사용자 음성을 발화한 사용자를 식별하는 단계(S1610)는 제1 사용자 음성이 입력되는 시간 구간 동안 영상에 포함된 사용자의 입모양에 기초하여 제1 사용자 음성을 발화한 사용자를 식별하고, 제2 응답 사운드를 제공하는 단계(S1640)는 제2 사용자 음성이 입력되는 시간 구간 동안 영상에 포함된 식별된 사용자의 입모양이 변경되면, 제1 사용자 음성을 발화한 사용자와 제2 사용자 음성을 발화한 사용자가 동일한 것으로 식별할 수 있다.

[269]

[270] 이상과 같은 본 개시의 다양한 실시 예에 따르면, 전자 장치는 복수의 사용자 각각으로부터 사용자 음성이 입력되면, 기존 사용자 음성에 기초하여 복수의 사용자 음성을 처리, 유보 또는 별도 처리함에 따라 컨텍스트를 유지한 채로 응답을 제공할 수 있다.

[271] 한편, 이상에서는 한 명의 사용자에 대하여 하나의 발화 히스토리 정보가 생성되는 것으로 설명하였으나, 이에 한정되는 것은 아니다. 예를 들어, 프로세서는 사용자 A의 제1 발화 히스토리 정보 및 사용자 B의 제2 발화 히스토리 정보가 저장된 상태에서 사용자 A의 제1 사용자 음성을 입력되면, 인공지능 모델에 제1 사용자 음성 및 제1 발화 히스토리 정보를 입력하여 획득된 제1 응답 사운드 및 인공지능 모델에 제1 사용자 음성 및 제2 발화 히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공할 수 있다. 이러한 동작을 통해 프로세서는 사용자에게 복수의 화자와 이야기하는 느낌을 제공할 수도 있다.

[272] 한편, 프로세서는 인터랙션이 완료된 발화 히스토리 정보에 대하여 주제, 진행 완료 정보, 진행 중인 정보, 시간 정보, 사용자 정보, 마지막 인터랙션 정보, 키워드 및 내용 요약 등의 정보를 획득하고, 이를 스토리지에 저장할 수 있다. 이후, 사용자가 과거 발화 히스토리 정보를 썬칭하는 사용자 음성을 발화하면, 프로세서는 스토리지에 저장된 정보에 기초하여 사용자 음성에 대응되는 발화 히스토리 정보를 획득할 수 있다. 그리고, 프로세서는 추가 사용자 음성이 입력되면, 획득된 발화 히스토리 정보에 기초하여 추가 사용자 음성에 대한 응답 사운드를 제공할 수 있다. 즉, 사용자는 과거 발화에 이어서 인터랙션을 수행할 수도 있다.

[273]

[274] 한편, 본 개시의 실시 예에 따르면, 이상에서 설명된 다양한 실시 예들은

기기(machine)(예: 컴퓨터)로 읽을 수 있는 저장 매체(machine-readable storage media)에 저장된 명령어를 포함하는 소프트웨어로 구현될 수 있다. 기기는, 저장 매체로부터 저장된 명령어를 호출하고, 호출된 명령어에 따라 동작이 가능한 장치로서, 개시된 실시 예들에 따른 전자 장치(예: 전자 장치(A))를 포함할 수 있다. 명령이 프로세서에 의해 실행될 경우, 프로세서가 직접, 또는 프로세서의 제어 하에 다른 구성요소들을 이용하여 명령에 해당하는 기능을 수행할 수 있다. 명령은 컴파일러 또는 인터프리터에 의해 생성 또는 실행되는 코드를 포함할 수 있다. 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 신호(signal)를 포함하지 않으며 실재(tangible)하다는 것을 의미할 뿐 데이터가 저장매체에 반영구적 또는 임시적으로 저장됨을 구분하지 않는다.

[275] 또한, 본 개시의 일 실시 예에 따르면, 이상에서 설명된 다양한 실시 예들에 따른 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다. 컴퓨터 프로그램 제품은 기기로 읽을 수 있는 저장 매체(예: compact disc read only memory (CD-ROM))의 형태로, 또는 어플리케이션 스토어(예: 플레이 스토어™)를 통해 온라인으로 배포될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

[276] 또한, 본 개시의 일 실시 예에 따르면, 이상에서 설명된 다양한 실시 예들은 소프트웨어(software), 하드웨어(hardware) 또는 이들의 조합을 이용하여 컴퓨터(computer) 또는 이와 유사한 장치로 읽을 수 있는 기록 매체 내에서 구현될 수 있다. 일부 경우에 있어 본 명세서에서 설명되는 실시 예들이 프로세서 자체로 구현될 수 있다. 소프트웨어적인 구현에 의하면, 본 명세서에서 설명되는 절차 및 기능과 같은 실시 예들은 별도의 소프트웨어 모듈들로 구현될 수 있다. 소프트웨어 모듈들 각각은 본 명세서에서 설명되는 하나 이상의 기능 및 동작을 수행할 수 있다.

[277] 한편, 상술한 다양한 실시 예들에 따른 기기의 프로세싱 동작을 수행하기 위한 컴퓨터 명령어(computer instructions)는 비일시적 컴퓨터 판독 가능 매체(non-transitory computer-readable medium)에 저장될 수 있다. 이러한 비일시적 컴퓨터 판독 가능 매체에 저장된 컴퓨터 명령어는 특정 기기의 프로세서에 의해 실행되었을 때 상술한 다양한 실시 예에 따른 기기에서의 처리 동작을 특정 기기가 수행하도록 한다. 비일시적 컴퓨터 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 비일시적 컴퓨터 판독 가능 매체의 구체적인 예로는, CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등이 있을 수 있다.

- [278] 또한, 상술한 다양한 실시 예들에 따른 구성 요소(예: 모듈 또는 프로그램) 각각은 단수 또는 복수의 개체로 구성될 수 있으며, 전술한 해당 서버 구성 요소들 중 일부 서버 구성 요소가 생략되거나, 또는 다른 서버 구성 요소가 다양한 실시 예에 더 포함될 수 있다. 대체적으로 또는 추가적으로, 일부 구성 요소들(예: 모듈 또는 프로그램)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성 요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 다양한 실시 예들에 따른, 모듈, 프로그램 또는 다른 구성 요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.
- [279] 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시에 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어져서는 안될 것이다.

청구범위

- [청구항 1] 전자 장치에 있어서,
 마이크;
 스피커; 및
 상기 마이크 및 상기 스피커와 전기적으로 연결된 프로세서;를 포함하며,
 상기 프로세서는,
 상기 마이크를 통해 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을
 발화한 사용자를 식별하고 인공지능 알고리즘을 통해 학습된 인공지능
 모델에 상기 제1 사용자 음성을 입력하여 획득된 제1 응답 사운드를 상기
 스피커를 통해 제공하며,
 상기 마이크를 통해 제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을
 발화한 사용자를 식별하고, 상기 제1 사용자 음성을 발화한 사용자와
 상기 제2 사용자 음성을 발화한 사용자가 동일하면, 상기 인공지능
 모델에 상기 제2 사용자 음성 및 발화 히스토리 정보를 입력하여 획득된
 제2 응답 사운드를 상기 스피커를 통해 제공하는, 전자 장치.
- [청구항 2] 제1항에 있어서,
 스토리지;를 더 포함하고,
 상기 프로세서는,
 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한
 사용자가 동일하지 않으면, 상기 제2 사용자 음성을 상기 스토리지에
 저장하고,
 상기 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자
 음성이 입력되지 않으면, 상기 인공지능 모델에 상기 제2 사용자 음성을
 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 제공하는, 전자
 장치.
- [청구항 3] 제1항에 있어서,
 상기 프로세서는,
 상기 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 제1
 사용자를 식별하고 상기 인공지능 모델에 상기 제1 사용자 음성을
 입력하여 획득된 상기 제1 응답 사운드를 상기 스피커를 통해 제공하며,
 상기 제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을 발화한 제2
 사용자를 식별하고, 상기 인공지능 모델에 상기 제2 사용자 음성을
 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 상기 제1 응답
 사운드와 구분되게 제공하는, 전자 장치.
- [청구항 4] 제3항에 있어서,
 상기 프로세서는,
 상기 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 상기 인공지능

모델에 상기 제1 추가 사용자 음성 및 상기 제1 사용자 음성에 대응되는 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를 상기 스피커를 통해 제공하고,

상기 제2 사용자에게 의해 제2 추가 사용자 음성이 입력되면, 상기 인공지능 모델에 상기 제2 추가 사용자 음성 및 상기 제2 사용자 음성에 대응되는 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 상기 스피커를 통해 상기 제1 추가 응답 사운드와 구분되게 제공하는, 전자 장치.

[청구항 5] 제4항에 있어서,
상기 프로세서와 전기적으로 연결된 디스플레이;를 더 포함하고,
상기 프로세서는,
상기 제1 추가 응답 사운드를 제공하며 상기 제1 사용자에게 대응되는 제1 UI를 상기 디스플레이를 통해 디스플레이하고,
상기 제2 추가 응답 사운드를 제공하며 상기 제2 사용자에게 대응되는 제2 UI를 상기 디스플레이를 통해 디스플레이하는, 전자 장치.

[청구항 6] 제1항에 있어서,
상기 프로세서는,
상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하지 않으면, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인을 획득하고,
상기 제1 도메인 및 상기 제2 도메인이 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 상기 발화 히스토리 정보를 입력하여 획득된 상기 제2 응답 사운드를 상기 스피커를 통해 제공하는, 전자 장치.

[청구항 7] 제1항에 있어서,
상기 프로세서와 전기적으로 연결된 스토리지;를 더 포함하고,
상기 프로세서는,
상기 제2 사용자 음성이 상기 제1 사용자 음성이 입력된 시점으로부터 기설정된 제1 시간 내에 입력되고, 상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일하며, 상기 제1 사용자 음성에 대응되는 제1 도메인 및 상기 제2 사용자 음성에 대응되는 제2 도메인이 서로 다르면, 상기 제1 응답 사운드의 제공을 생략하고 상기 제1 사용자 음성을 상기 스토리지에 저장하며, 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답 사운드를 상기 스피커를 통해 제공하는, 전자 장치.

[청구항 8] 제7항에 있어서,
상기 프로세서는,
상기 사용자로부터 상기 제2 도메인에 대응되는 사용자 음성이 기설정된 제2 시간 이상 입력되지 않으면, 상기 인공지능 모델에 상기 스토리지에

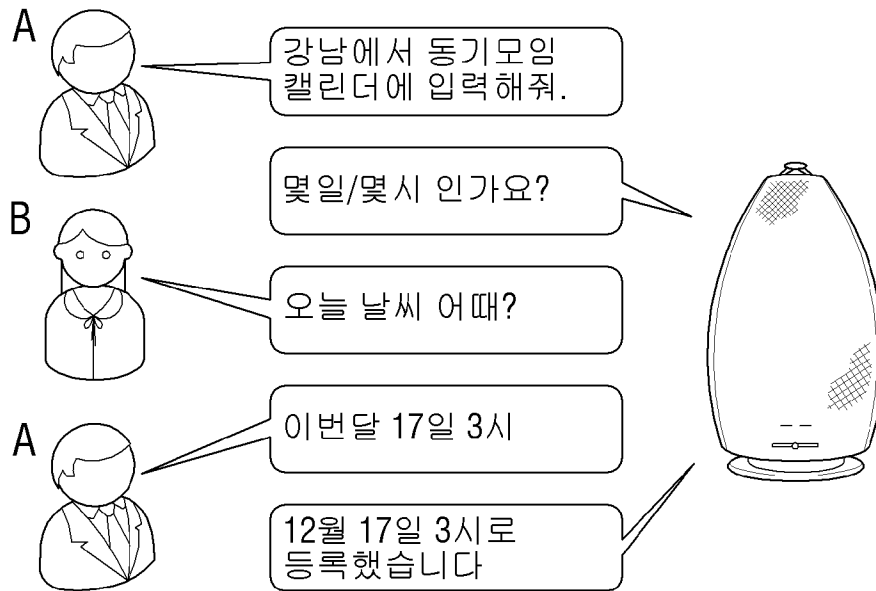
- 저장된 상기 제1 사용자 음성을 입력하여 획득된 상기 제1 응답 사운드를
상기 스피커를 통해 제공하는, 전자 장치.
- [청구항 9] 제1항에 있어서,
카메라;를 더 포함하고,
상기 프로세서는,
상기 카메라에 의해 촬영된 영상 및 상기 마이크를 통해 입력된 사용자
음성 중 적어도 하나에 기초하여 상기 사용자를 식별하는, 전자 장치.
- [청구항 10] 제9항에 있어서,
상기 프로세서는,
상기 제1 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된
사용자의 입모양에 기초하여 상기 제1 사용자 음성을 발화한 사용자를
식별하고,
상기 제2 사용자 음성이 입력되는 시간 구간 동안 상기 영상에 포함된
상기 식별된 사용자의 입모양이 변경되면, 상기 제1 사용자 음성을
발화한 사용자와 상기 제2 사용자 음성을 발화한 사용자가 동일한 것으로
식별하는, 전자 장치.
- [청구항 11] 전자 장치의 제어 방법에 있어서,
제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 사용자를
식별하는 단계;
인공지능 알고리즘을 통해 학습된 인공지능 모델에 상기 제1 사용자
음성을 입력하여 획득된 제1 응답 사운드를 제공하는 단계;
제2 사용자 음성이 입력되면, 상기 제2 사용자 음성을 발화한 사용자를
식별하는 단계; 및
상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한
사용자가 동일하면, 상기 인공지능 모델에 상기 제2 사용자 음성 및 발화
히스토리 정보를 입력하여 획득된 제2 응답 사운드를 제공하는 단계;를
포함하는, 제어 방법.
- [청구항 12] 제11항에 있어서,
상기 제1 사용자 음성을 발화한 사용자와 상기 제2 사용자 음성을 발화한
사용자가 동일하지 않으면, 상기 제2 사용자 음성을 저장하는 단계; 및
상기 제1 사용자 음성을 발화한 사용자로부터 기설정된 시간 이상 사용자
음성이 입력되지 않으면, 상기 인공지능 모델에 상기 제2 사용자 음성을
입력하여 획득된 제3 응답 사운드를 제공하는 단계;를 더 포함하는, 제어
방법.
- [청구항 13] 제11항에 있어서,
상기 제1 사용자 음성을 발화한 사용자를 식별하는 단계는,
상기 제1 사용자 음성이 입력되면, 상기 제1 사용자 음성을 발화한 제1
사용자를 식별하고,

상기 제1 응답 사운드를 제공하는 단계는,
 상기 인공지능 모델에 상기 제1 사용자 음성을 입력하여 획득된 상기 제1
 응답 사운드를 제공하며,
 상기 제2 사용자 음성을 발화한 사용자를 식별하는 단계는,
 상기 제2 사용자 음성을 발화한 제2 사용자를 식별하고,
 상기 제어 방법은,
 상기 인공지능 모델에 상기 제2 사용자 음성을 입력하여 획득된 제3 응답
 사운드를 상기 제1 응답 사운드와 구분되게 제공하는 단계;를 더
 포함하는, 제어 방법.

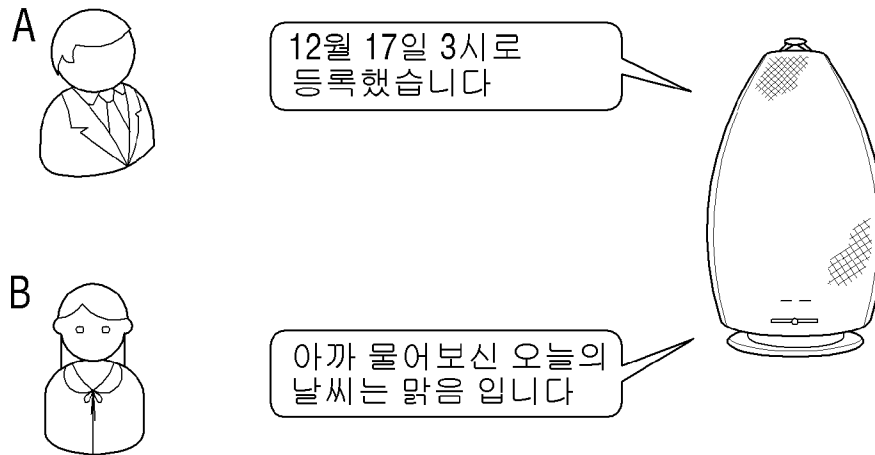
[청구항 14] 제13항에 있어서,
 상기 제1 사용자에 의해 제1 추가 사용자 음성이 입력되면, 상기 인공지능
 모델에 상기 제1 추가 사용자 음성 및 상기 제1 사용자 음성에 대응되는
 제1 발화 히스토리 정보를 입력하여 획득된 제1 추가 응답 사운드를
 제공하는 단계; 및
 상기 제2 사용자에 의해 제2 추가 사용자 음성이 입력되면, 상기 인공지능
 모델에 상기 제2 추가 사용자 음성 및 상기 제2 사용자 음성에 대응되는
 제2 발화 히스토리 정보를 입력하여 획득된 제2 추가 응답 사운드를 상기
 제1 추가 응답 사운드와 구분되게 제공하는 단계;를 더 포함하는, 제어
 방법.

[청구항 15] 제14항에 있어서,
 상기 제1 추가 응답 사운드를 제공하는 단계는,
 상기 제1 추가 응답 사운드를 제공하며 상기 제1 사용자에 대응되는 제1
 UI를 디스플레이하고,
 상기 제2 추가 응답 사운드를 상기 제1 추가 응답 사운드와 구분되게
 제공하는 단계는,
 상기 제2 추가 응답 사운드를 제공하며 상기 제2 사용자에 대응되는 제2
 UI를 디스플레이하는, 제어 방법.

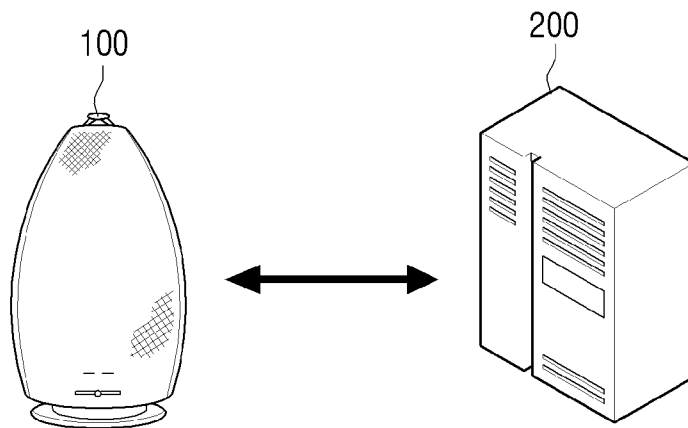
[도 1a]



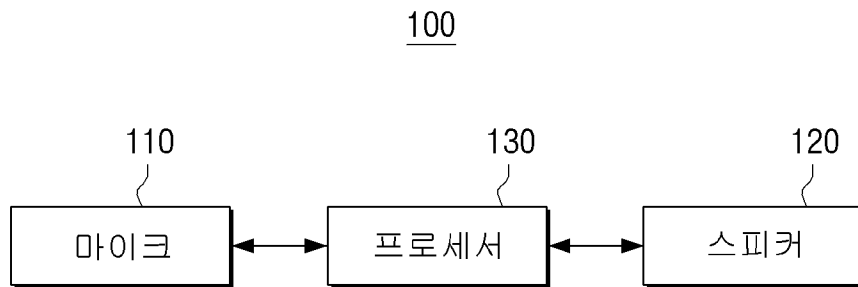
[도 1b]



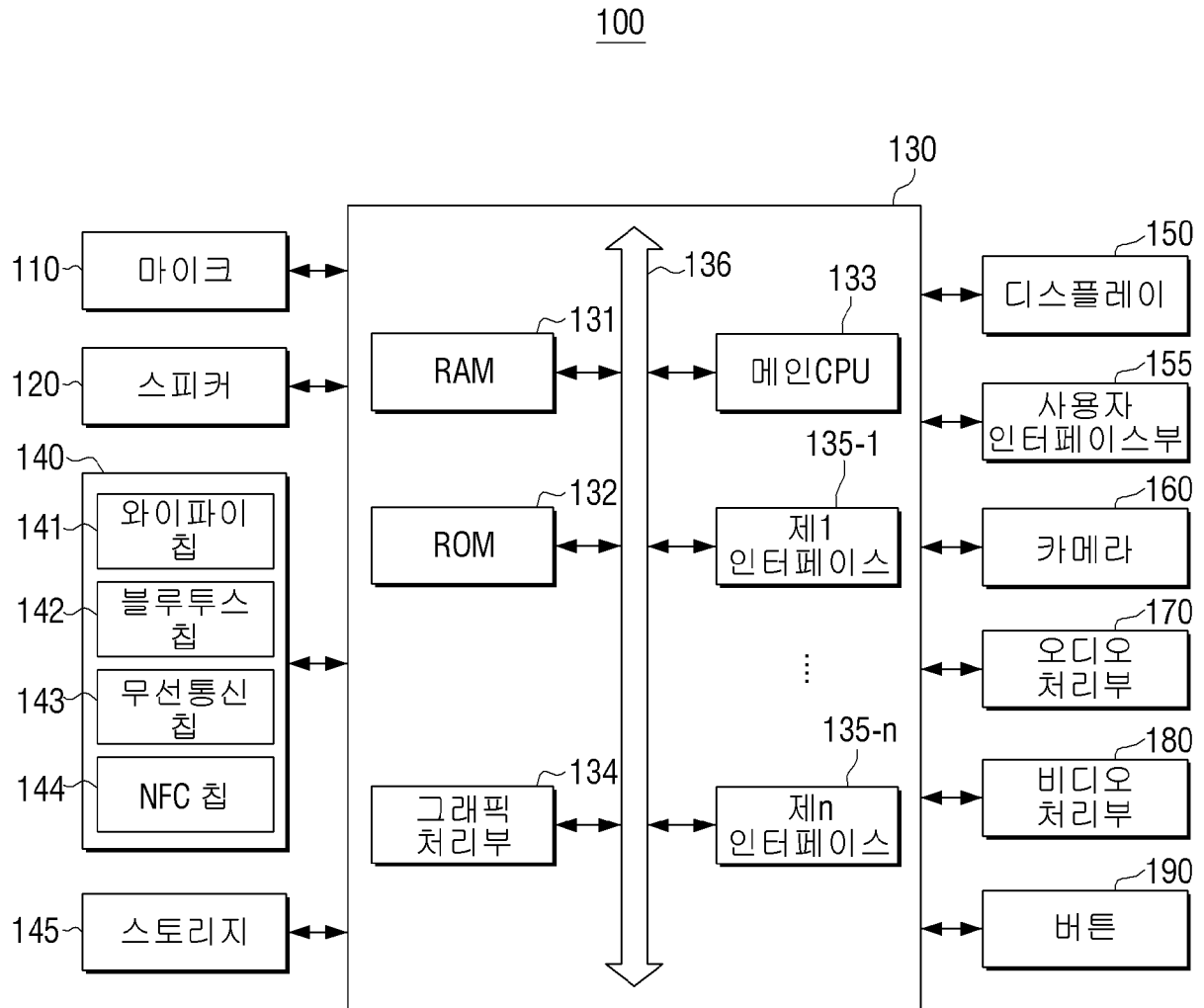
[도 2]



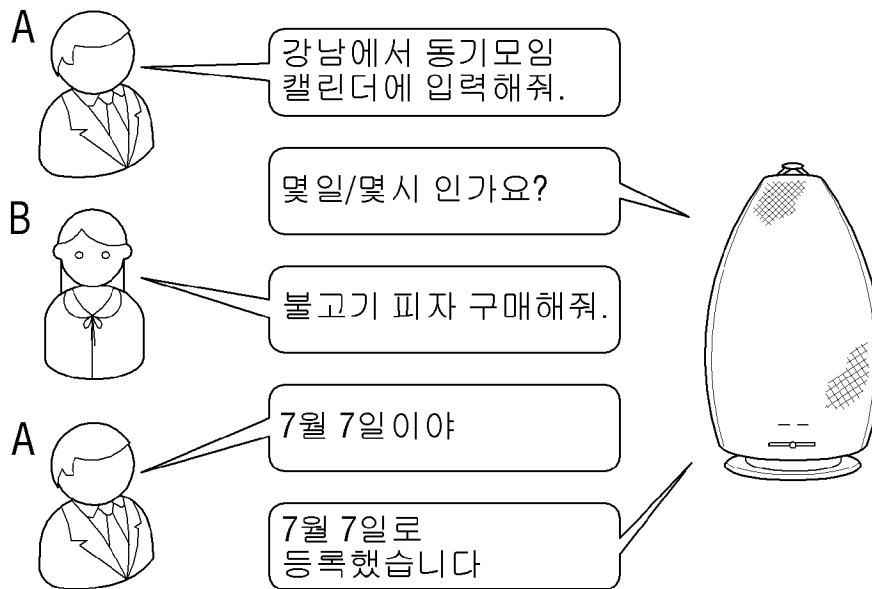
[도3a]



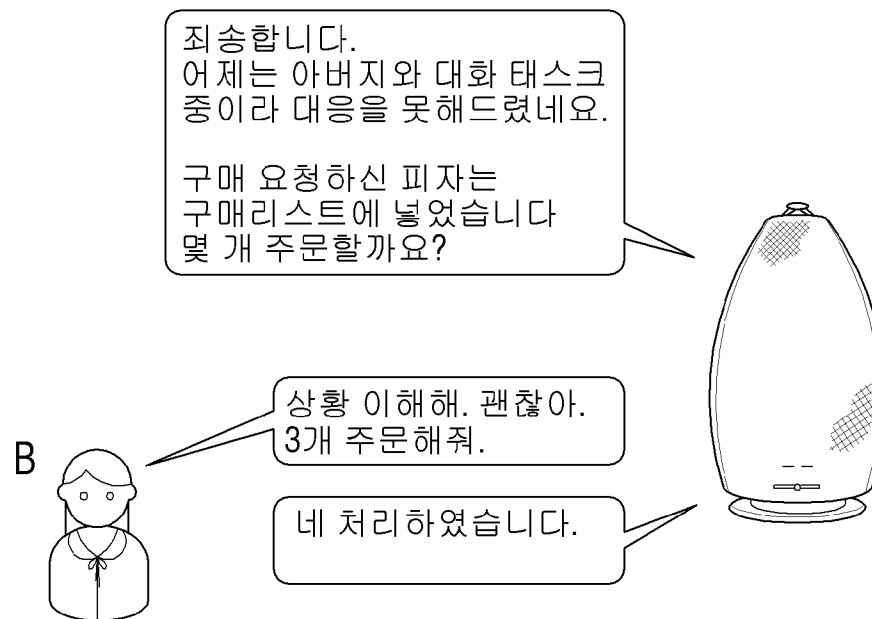
[도3b]



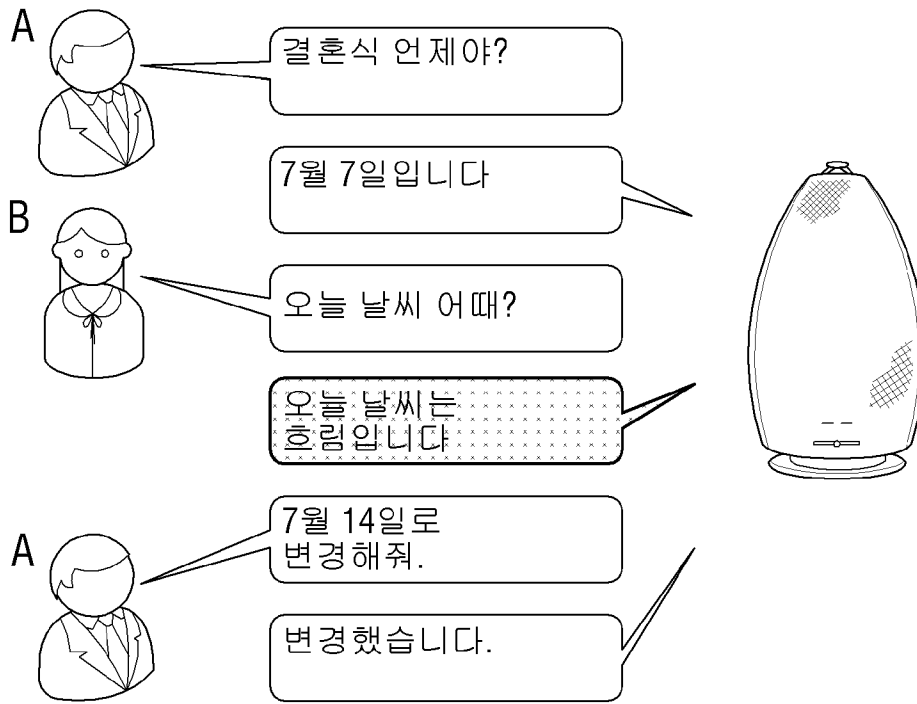
[도4a]



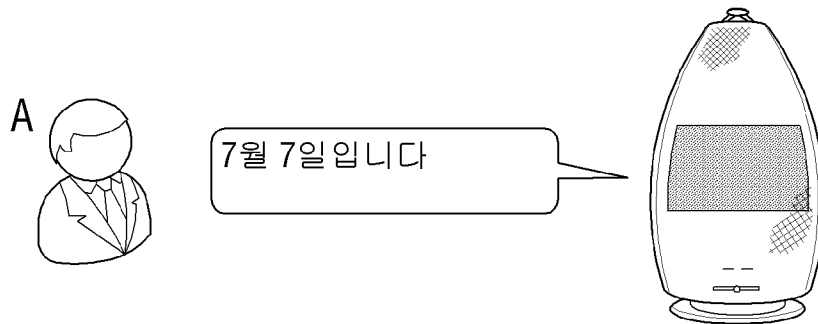
[도4b]



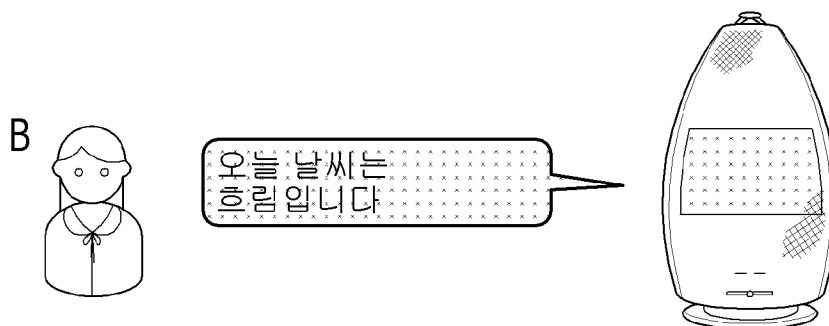
[도5a]



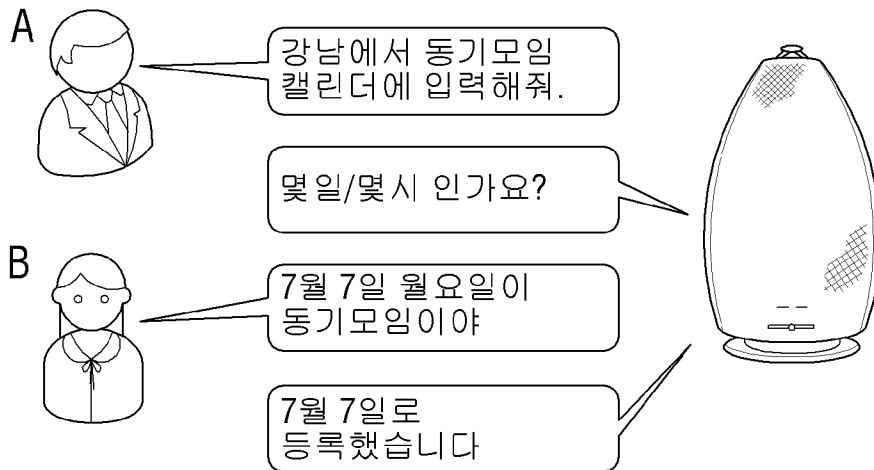
[도5b]



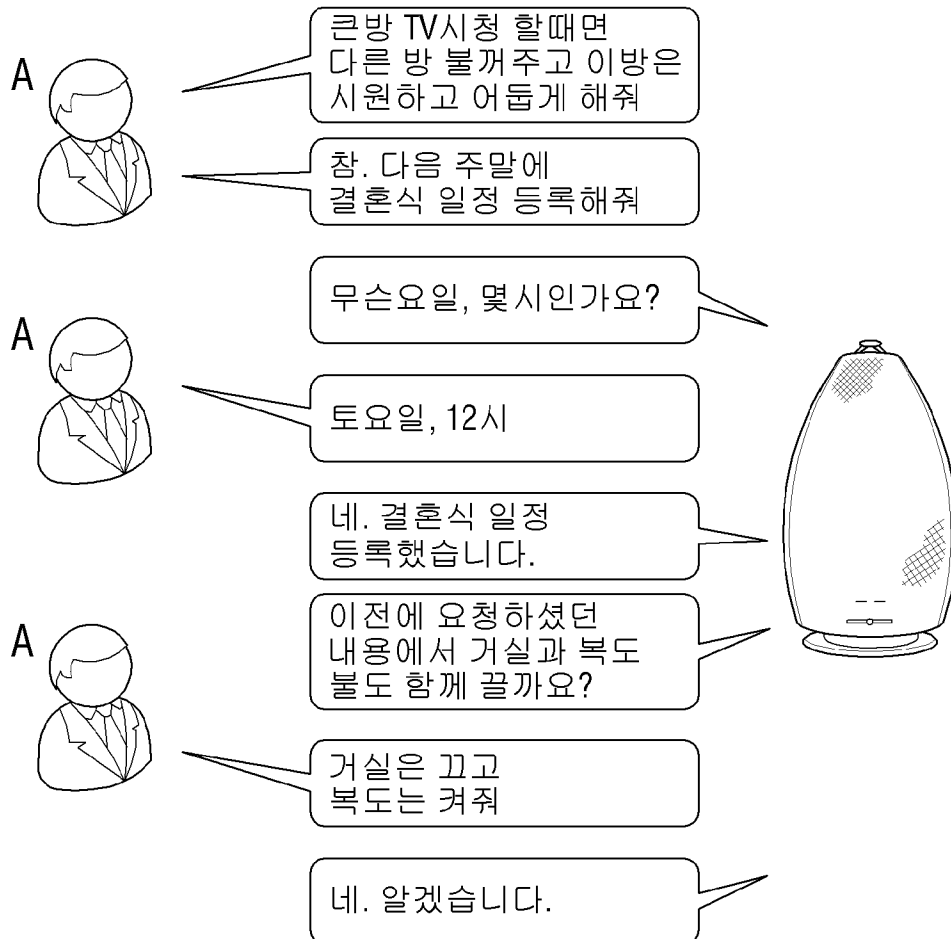
[도5c]



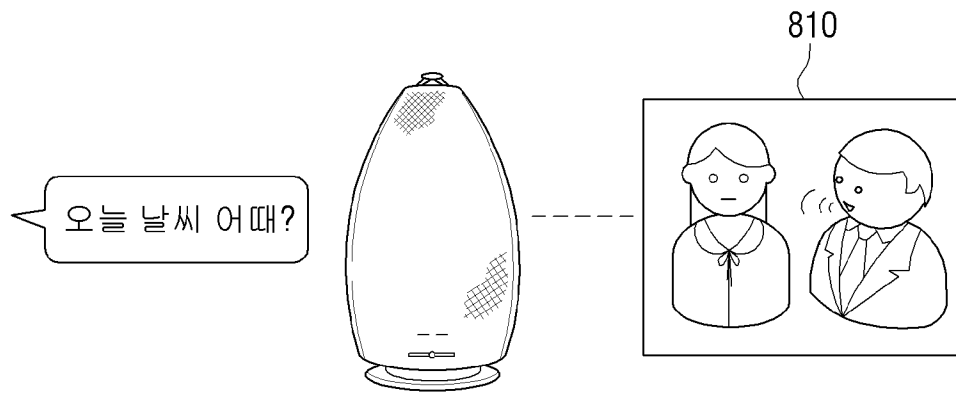
[도6]



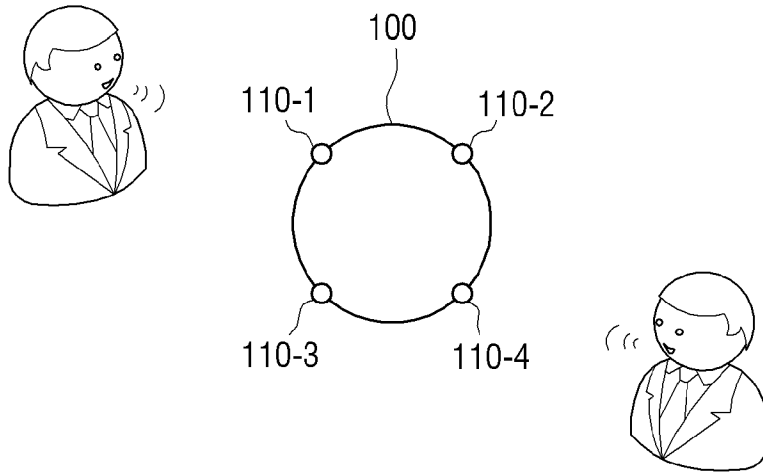
[도7]



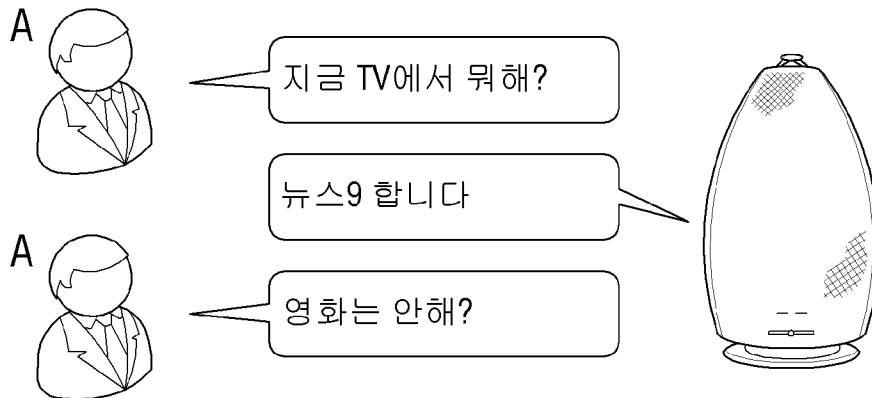
[도8a]



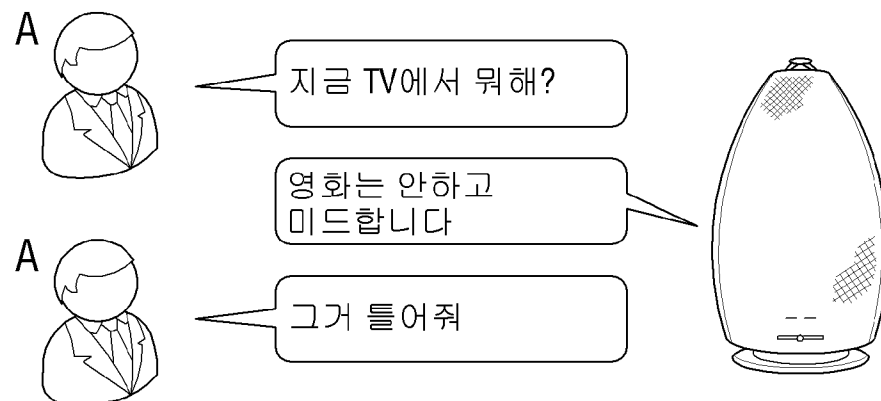
[도8b]



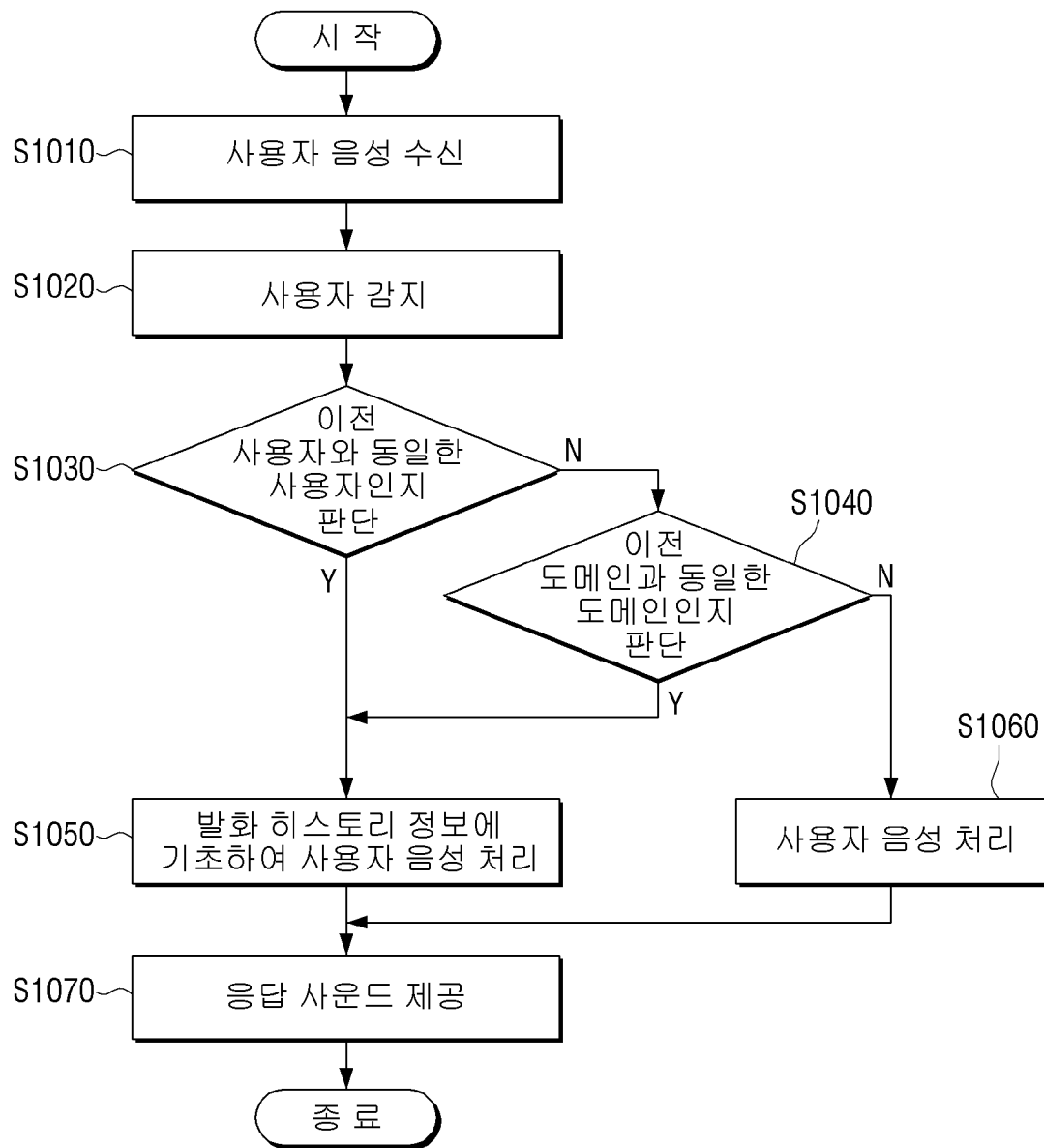
[도9a]



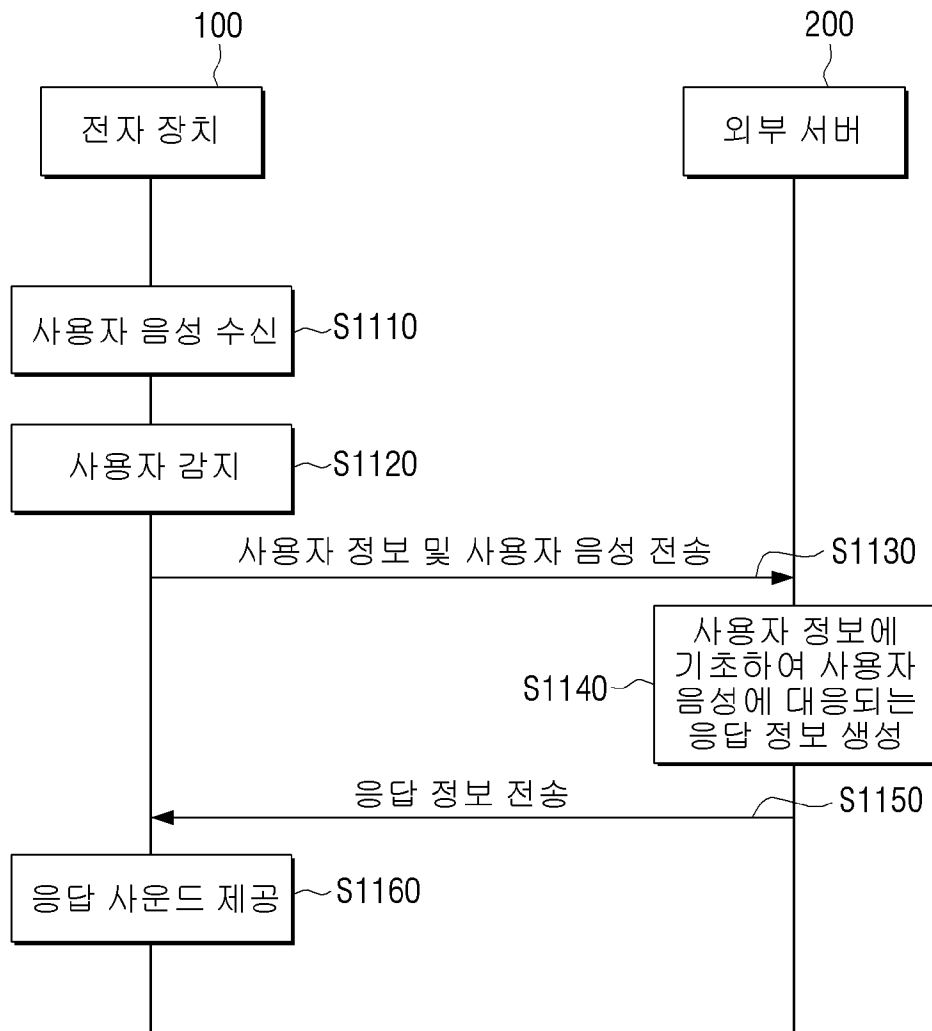
[도9b]



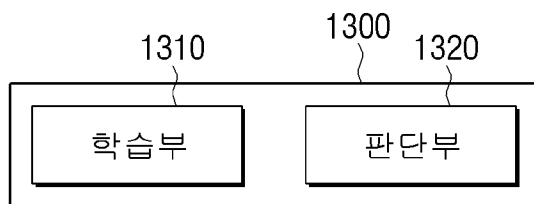
[도10]



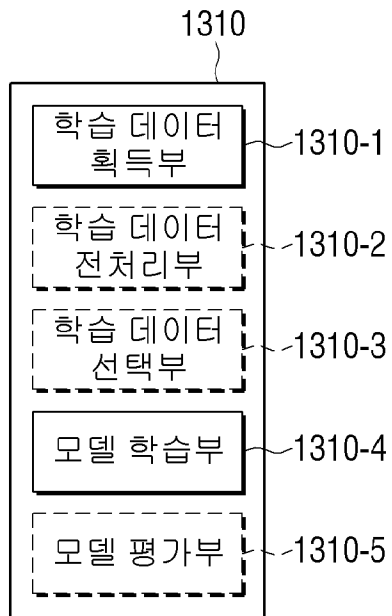
[도11]



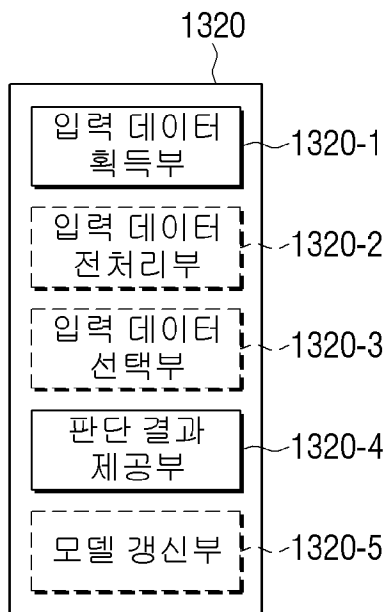
[도12]



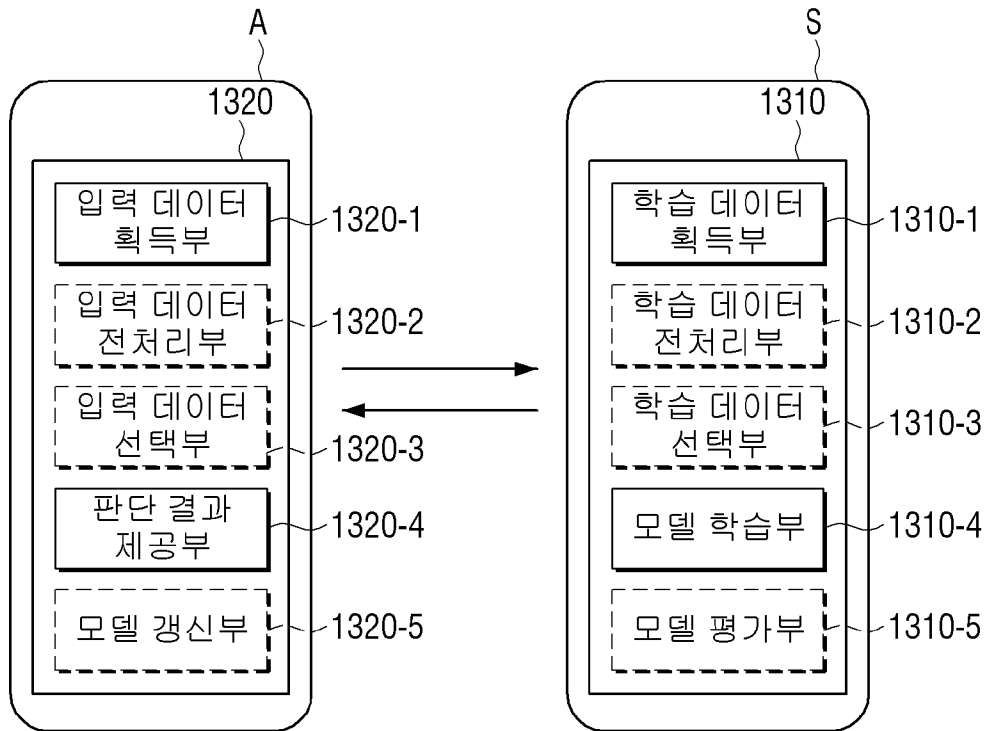
[도13]



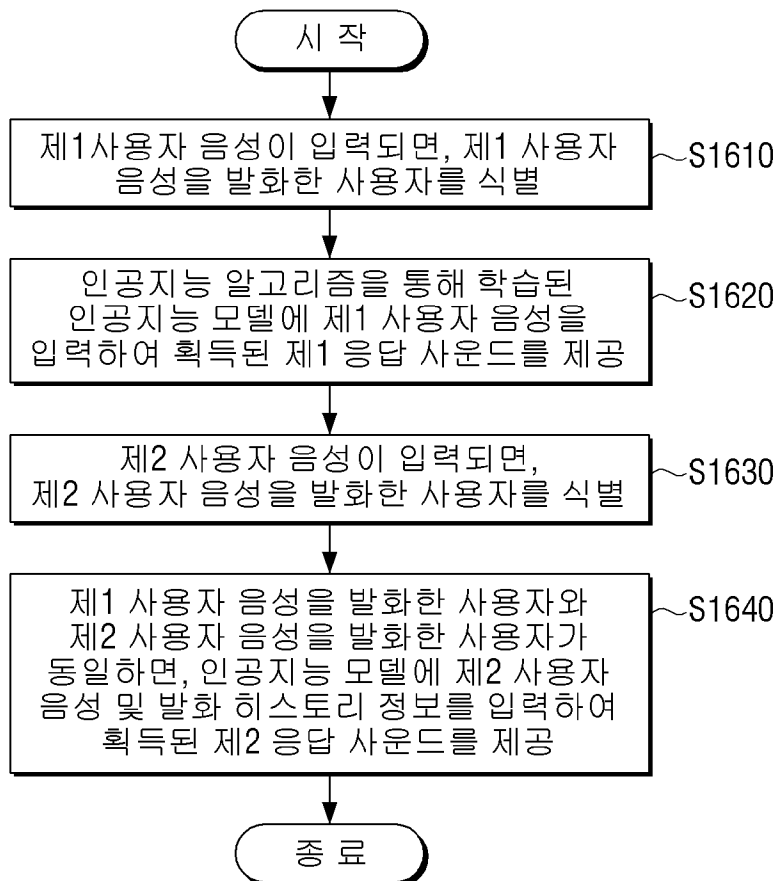
[도14]



[도15]



[도16]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/000036

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/16(2006.01)i, G06F 3/00(2006.01)i, G06F 3/01(2006.01)i, G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 3/16; G06F 17/20; G06F 17/30; G10L 15/02; G10L 15/22; G10L 15/26; G10L 15/28; G06F 3/00; G06F 3/01; G06K 9/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: multi-user, identical, voice, history, query, response

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 9098467 B1 (RAWLES LLC.) 04 August 2015 See column 4, line 5-column 8, line 16; and figures 1-3.	1,3-5,9,11,13-15
A		2,6-8,10,12
Y	KR 10-2016-0065671 A (SAMSUNG ELECTRONICS CO., LTD.) 09 June 2016 See paragraphs [0075]-[0234]; claims 1-10; and figures 2-18.	1,3-5,9,11,13-15
A	US 9424840 B1 (RAWLES LLC.) 23 August 2016 See column 8, line 41-column 9, line 23; and figure 1A.	1-15
A	KR 10-2014-0039868 A (KUM YOUNG CO., LTD.) 02 April 2014 See paragraphs [0082]-[0143]; and figures 4-6.	1-15
A	KR 10-2014-0068752 A (LG ELECTRONICS INC.) 09 June 2014 See paragraphs [0056]-[0171]; and figures 2-9.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

01 APRIL 2019 (01.04.2019)

Date of mailing of the international search report

02 APRIL 2019 (02.04.2019)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsu-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2019/000036

Patent document cited in search report	Publication date	Patent family member	Publication date
US 9098467 B1	04/08/2015	None	
KR 10-2016-0065671 A	09/06/2016	EP 3227800 A1 US 2016-0154777 A1 WO 2016-089079 A1	11/10/2017 02/06/2016 09/06/2016
US 9424840 B1	23/08/2016	US 10026394 B1 US 9299346 B1	17/07/2018 29/03/2016
KR 10-2014-0039868 A	02/04/2014	KR 10-1429138 B1	11/08/2014
KR 10-2014-0068752 A	09/06/2014	KR 10-1556173 B1 US 2014-0149118 A1 WO 2014-084413 A1	30/09/2015 29/05/2014 05/06/2014

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 3/16(2006.01)i, G06F 3/00(2006.01)i, G06F 3/01(2006.01)i, G06K 9/00(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 3/16; G06F 17/20; G06F 17/30; G10L 15/02; G10L 15/22; G10L 15/26; G10L 15/28; G06F 3/00; G06F 3/01; G06K 9/00

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 다중 사용자, 동일, 음성, 히스토리, 질의, 응답

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	US 9098467 B1 (RAWLES LLC) 2015.08.04 컬럼 4, 라인 5 - 컬럼 8, 라인 16; 및 도면 1-3 참조.	1,3-5,9,11,13-15
A		2,6-8,10,12
Y	KR 10-2016-0065671 A (삼성전자주식회사) 2016.06.09 단락 [0075]-[0234]; 청구항 1-10; 및 도면 2-18 참조.	1,3-5,9,11,13-15
A	US 9424840 B1 (RAWLES LLC) 2016.08.23 컬럼 8, 라인 41 - 컬럼 9, 라인 23; 및 도면 1A 참조.	1-15
A	KR 10-2014-0039868 A (주식회사 금영) 2014.04.02 단락 [0082]-[0143]; 및 도면 4-6 참조.	1-15
A	KR 10-2014-0068752 A (엘지전자 주식회사) 2014.06.09 단락 [0056]-[0171]; 및 도면 2-9 참조.	1-15

☐ 추가 문헌이 C(계속)에 기재되어 있습니다.

☒ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

“A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2019년 04월 01일 (01.04.2019)

국제조사보고서 발송일

2019년 04월 02일 (02.04.2019)

ISA/KR의 명칭 및 우편주소



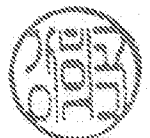
대한민국 특허청
(35208) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 +82-42-481-8578

심사관

강희국

전화번호 +82-42-481-8264



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
US 9098467 B1	2015/08/04	없음	
KR 10-2016-0065671 A	2016/06/09	EP 3227800 A1 US 2016-0154777 A1 WO 2016-089079 A1	2017/10/11 2016/06/02 2016/06/09
US 9424840 B1	2016/08/23	US 10026394 B1 US 9299346 B1	2018/07/17 2016/03/29
KR 10-2014-0039868 A	2014/04/02	KR 10-1429138 B1	2014/08/11
KR 10-2014-0068752 A	2014/06/09	KR 10-1556173 B1 US 2014-0149118 A1 WO 2014-084413 A1	2015/09/30 2014/05/29 2014/06/05