



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0098250
(43) 공개일자 2021년08월10일

- | | |
|---|--|
| (51) 국제특허분류(Int. Cl.)
G06F 3/16 (2018.01) G06F 40/20 (2020.01)
G06N 3/08 (2006.01) G10L 15/183 (2013.01)
G10L 17/26 (2013.01)
(52) CPC특허분류
G06F 3/167 (2013.01)
G06F 40/20 (2020.01)
(21) 출원번호 10-2020-0012200
(22) 출원일자 2020년01월31일
심사청구일자 없음 | (71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
(72) 발명자
윤정현
경기도 수원시 영통구 삼성로 129(매탄동)
(74) 대리인
정홍식, 김태현 |
|---|--|

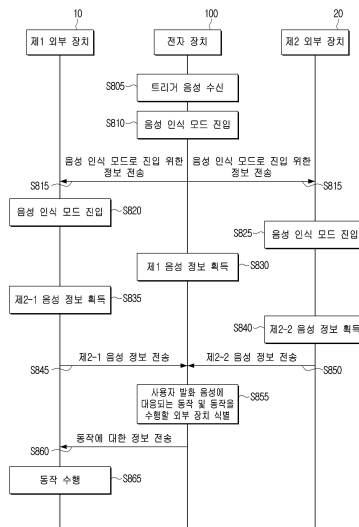
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 전자 장치 및 이의 제어 방법

(57) 요약

전자 장치 및 그 제어 방법이 개시된다. 본 전자 장치의 제어 방법은, 사용자의 트리거 음성을 수신하는 단계; 상기 트리거 음성에 응답하여 음성 인식 모드로 진입하고, 음성 인식 모드로 진입하기 위한 정보를 맥내에 위치한 적어도 하나의 외부 장치로 전송하는 단계; 음성 인식 모드로 동작하는 동안 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득하고, 상기 적어도 하나의 외부 장치로부터 상기 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득하는 단계; 상기 제1 음성 정보 및 상기 적어도 하나의 제2 음성 정보를 바탕으로, 상기 사용자가 발화한 음성에 대응되는 동작 및 상기 동작을 수행할 외부 장치를 식별하는 단계; 및 상기 식별된 외부 장치로 상기 동작에 대한 정보를 전송하는 단계;를 포함한다.

대표도 - 도8



(52) CPC특허분류

- G06N 3/08* (2013.01)
 - G10L 15/183* (2013.01)
 - G10L 15/22* (2013.01)
 - G10L 15/26* (2013.01)
 - G10L 17/26* (2013.01)
-

명세서

청구범위

청구항 1

전자 장치의 제어 방법에 있어서,

사용자의 트리거 음성을 수신하는 단계;

상기 트리거 음성에 응답하여 상기 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 음성 인식 모드로 진입하기 위한 정보를 맥내에 위치한 적어도 하나의 외부 장치로 전송하는 단계;

음성 인식 모드로 동작하는 동안 상기 전자 장치에 포함된 마이크로로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득하고, 상기 적어도 하나의 외부 장치로부터 상기 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득하는 단계;

상기 제1 음성 정보 및 상기 적어도 하나의 제2 음성 정보를 바탕으로, 상기 사용자가 발화한 음성에 대응되는 동작 및 상기 동작을 수행할 외부 장치를 식별하는 단계; 및

상기 식별된 외부 장치로 상기 동작에 대한 정보를 전송하는 단계;를 포함하는 제어 방법.

청구항 2

제1항에 있어서,

상기 식별하는 단계는,

상기 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 ASR(Automatic Speech Recognition) 모듈에 입력하여 상기 제1 음성 정보에 대응되는 제1 텍스트 정보 및 적어도 하나의 제2 음성 정보에 대응되는 적어도 하나의 제2 텍스트 정보를 획득하는 단계; 및

상기 제1 텍스트 정보 및 상기 적어도 하나의 제2 텍스트 정보를 바탕으로 상기 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 획득하는 단계;를 더 포함하는 제어 방법.

청구항 3

제2항에 있어서,

상기 최종 텍스트 정보를 획득하는 단계는,

상기 제1 텍스트 정보와 상기 제2 텍스트 정보가 동일한 경우, 상기 제1 텍스트 정보를 상기 최종 텍스트 정보로 식별하고,

상기 제1 텍스트 정보와 상기 제2 텍스트 정보가 상이한 경우, 상기 제1 텍스트 정보 및 상기 제2 텍스트 정보를 인공 신경망 모델에 입력하여, 최종 텍스트 정보를 획득하는 것을 특징으로 하는 제어 방법.

청구항 4

제2 항에 있어서,

상기 최종 텍스트 정보를 NLU(Natural Language Understanding) 모듈에 입력하여, 상기 사용자가 발화한 음성 에 대응되는 동작에 대한 정보를 획득하는 단계;를 더 포함하는 제어 방법.

청구항 5

제2항에 있어서,

상기 최종 텍스트 정보는 적어도 하나의 키워드를 포함하고,

상기 식별하는 단계는,

상기 적어도 하나의 키워드를 바탕으로 상기 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 단계;를 더 포함하는 제어 방법.

청구항 6

제3항에 있어서,

상기 최종 텍스트 정보는 상기 사용자가 발화한 음성의 음절 각각에 대응되는 복수의 키워드를 포함하고,

상기 식별하는 단계는,

상기 복수의 키워드를 바탕으로 상기 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 단계;를 더 포함하는 제어 방법.

청구항 7

제5항에 있어서,

상기 외부 장치를 식별하는 단계는,

상기 최종 텍스트 정보에 포함된 적어도 하나의 키워드 각각에 부여된 가중치를 통해 동작을 수행할 외부 장치가 식별되는 것을 특징으로 하는 제어 방법.

청구항 8

제1항에 있어서,

상기 적어도 하나의 외부 장치로 전송하는 단계는,

상기 정보를 수신한 적어도 하나의 외부 장치 중 영상 또는 오디오가 재생 중인 전자 장치를 음소거 모드로 진입하기 위한 정보를 더 전송하는 제어 방법.

청구항 9

제1항에 있어서,

상기 전자 장치 및 상기 적어도 하나의 외부 장치에 포함된 음성 인식 엔진은 동일한 것을 특징으로 하는 제어 방법.

청구항 10

전자 장치에 있어서,

통신 인터페이스;

입력 인터페이스;

적어도 하나의 인스트럭션을 저장한 메모리; 및

상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행하여 상기 전자 장치를 제어하는 프로세서를 포함하고,
상기 프로세서는,

상기 입력 인터페이스를 통해 사용자의 트리거 음성을 수신하고,

상기 트리거 음성에 응답하여 상기 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 상기 통신 인터페이스를 통해 음성 인식 모드로 진입하기 위한 정보를 맥내에 위치한 적어도 하나의 외부 장치로 전송하고,

음성 인식 모드로 동작하는 동안 상기 입력 인터페이스에 포함된 마이크로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득하고, 상기 통신 인터페이스를 통해 상기 적어도 하나의 외부 장치로부터 상기 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득하고,

상기 제1 음성 정보 및 상기 적어도 하나의 제2 음성 정보를 바탕으로, 상기 사용자가 발화한 음성에 대응되는 동작 및 상기 동작을 수행할 외부 장치를 식별하고,

상기 통신 인터페이스를 통해 상기 식별된 외부 장치로 상기 동작에 대한 정보를 전송하는 전자 장치.

청구항 11

제10항에 있어서,

상기 프로세서는,

상기 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 ASR 모듈에 입력하여 상기 제1 음성 정보에 대응되는 제1 텍스트 정보 및 적어도 하나의 제2 음성 정보에 대응되는 적어도 하나의 제2 텍스트 정보를 획득하고,

상기 제1 텍스트 정보 및 상기 적어도 하나의 제2 텍스트 정보를 바탕으로 상기 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 획득하는 전자 장치.

청구항 12

제11항에 있어서,

상기 프로세서는,

상기 제1 텍스트 정보와 상기 제2 텍스트 정보가 동일한 경우, 상기 제1 텍스트 정보를 상기 최종 텍스트 정보로 식별하고,

상기 제1 텍스트 정보와 상기 제2 텍스트 정보가 상이한 경우, 상기 제1 텍스트 정보 및 상기 제2 텍스트 정보를 인공 신경망 모델에 입력하여, 최종 텍스트 정보를 획득하는 것을 특징으로 하는 전자 장치.

청구항 13

제11 항에 있어서,

상기 프로세서는,

상기 최종 텍스트 정보를 NLU 모듈에 입력하여, 상기 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 획득하는 전자 장치.

청구항 14

제11항에 있어서,

상기 최종 텍스트 정보는 적어도 하나의 키워드를 포함하고,

상기 프로세서는,

상기 적어도 하나의 키워드를 바탕으로 상기 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 전자 장치.

청구항 15

제11항에 있어서,

상기 최종 텍스트 정보는 상기 사용자가 발화한 음성의 음절 각각에 대응되는 복수의 키워드를 포함하고,

상기 프로세서는,

상기 복수의 키워드를 바탕으로 상기 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 전자 장치.

청구항 16

제14항에 있어서,

상기 프로세서는,

상기 최종 텍스트 정보에 포함된 적어도 하나의 키워드 각각에 부여된 가중치를 통해 동작을 수행할 외부 장치가 식별되는 것을 특징으로 하는 전자 장치.

청구항 17

제10항에 있어서,

상기 프로세서는,

상기 정보를 수신한 적어도 하나의 외부 장치 중 영상 또는 오디오가 재생 중인 전자 장치를 음소거 모드로 진입하기 위한 정보를 더 전송하는 전자 장치

청구항 18

제10항에 있어서,

상기 전자 장치 및 상기 적어도 하나의 외부 장치에 포함된 음성 인식 엔진은 동일한 것을 특징으로 하는 전자 장치.

발명의 설명

기술 분야

[0001] 본 개시는 전자 장치 및 이의 제어 방법에 관한 것으로, 더욱 상세하게는 사용자의 트리거 음성에 응답하여, 외부 장치로 음성 인식 모드로 진입하기 위한 정보를 전송하는 전자 장치 및 이의 제어 방법에 관한 것이다.

배경 기술

[0002] 근래에는 인공지능 시스템이 다양한 분야에서 이용되고 있다. 인공지능 시스템은 기존의 룰(rule) 기반 스마트 시스템과 달리 기계가 스스로 학습하고 판단하며 똑똑해지는 시스템이다. 인공지능 시스템은 사용할수록 인식이 향상되고 사용자 취향을 보다 정확하게 이해할 수 있게 되어, 기존 룰 기반 스마트 시스템은 점차 딥러닝 기반의 인공지능 시스템으로 대체되고 있다.

[0003] 한편, 근래에는 전자 장치는 사용자 음성을 이용하여 다양한 동작(task)을 수행할 수 있다. 예를 들어, 전자 장

치는 사용자 음성을 통해 사용자의 음성에 대응되는 전자 장치의 동작을 수행하거나, 외부 기기를 제어하기 위한 동작을 수행할 수 있다.

[0004] 다만, 전자 장치는 사용자 음성을 통해 다양한 동작을 수행할 수 있으나, 태내에 사용자 음성을 인식할 수 있는 복수의 외부 장치가 있는 경우, 사용자의 음성 인식 모드로 진입하기 위한 트리거 음성이 수신된 외부 장치만 음성 인식 모드로 진입하게 되어, 인식률이 떨어지는 문제점이 존재하였다. 또한, 태내에 위치한 복수의 외부 장치가 사용자의 트리거 음성을 인식한 경우, 사용자가 의도하지 않은 외부 장치가 음성 인식 모드로 진입하게 되는 문제점이 존재하였다.

[0005] 또한, 태내의 복수의 외부 장치가 사용자의 트리거 음성을 인식하더라도, 미디어가 재생 중인 다른 외부 장치가 존재하거나, 사용자로부터 멀리 떨어진 외부 장치의 경우, 사용자의 발화 음성에 대한 인식률이 떨어지는 문제점이 존재하였다.

발명의 내용

해결하려는 과제

[0006] 본 개시는 상술한 문제점을 해결하기 위한 것으로, 본 개시의 목적은 사용자의 트리거 음성에 응답하여, 음성 인식모드로 진입하기 위한 정보를 외부 장치로 전송하는 전자 장치 및 이의 제어 방법에 관한 것이다.

과제의 해결 수단

[0007] 상술한 목적을 달성하기 위한 본 개시의 일 실시 예에 따른, 전자 장치의 제어 방법은, 사용자의 트리거 음성을 수신하는 단계; 상기 트리거 음성에 응답하여 상기 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 음성 인식 모드로 진입하기 위한 정보를 태내에 위치한 적어도 하나의 외부 장치로 전송하는 단계; 음성 인식 모드로 동작하는 동안 상기 전자 장치에 포함된 마이크로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보음성 정보를 획득하고, 상기 적어도 하나의 외부 장치로부터 상기 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보음성 정보를 획득하는 단계; 상기 제1 음성 정보음성 정보 및 상기 적어도 하나의 제2 음성 정보음성 정보를 바탕으로, 상기 사용자가 발화한 음성에 대응되는 동작 및 상기 동작을 수행할 외부 장치를 식별하는 단계; 및 상기 식별된 외부 장치로 상기 동작에 대한 정보를 전송하는 단계;를 포함한다.

[0008] 한편, 본 개시의 일 실시 예에 따른, 전자 장치는, 통신 인터페이스; 입력 인터페이스; 적어도 하나의 인스트럭션을 저장한 메모리; 및 상기 메모리에 저장된 적어도 하나의 인스트럭션을 실행하여 상기 전자 장치를 제어하는 프로세서를 포함하고, 상기 프로세서는, 상기 입력 인터페이스를 통해 사용자의 트리거 음성을 수신하고, 상기 트리거 음성에 응답하여 상기 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 상기 통신 인터페이스를 통해 음성 인식 모드로 진입하기 위한 정보를 태내에 위치한 적어도 하나의 외부 장치로 전송하고, 음성 인식 모드로 동작하는 동안 상기 입력 인터페이스에 포함된 마이크로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보음성 정보를 획득하고, 상기 통신 인터페이스를 통해 상기 적어도 하나의 외부 장치로부터 상기 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보음성 정보를 획득하고, 상기 제1 음성 정보음성 정보 및 상기 적어도 하나의 제2 음성 정보음성 정보를 바탕으로, 상기 사용자가 발화한 음성에 대응되는 동작 및 상기 동작을 수행할 외부 장치를 식별하고, 상기 통신 인터페이스를 통해 상기 식별된 외부 장치로 상기 동작에 대한 정보를 전송한다.

발명의 효과

[0009] 본 개시의 다양한 실시 예에 따르면, 트리거 음성이 수신되면, 전자 장치가 전자 태내에 위치한 모든 외부 장치를 음성 인식 모드로 진입하도록 제어함으로써, 사용자가 의도하지 않은 외부 장치만 음성 인식 모드로 진입하게 되는 문제점이 방지될 수 있다.

[0010] 또한, 적어도 하나의 외부 장치로부터 수신한 음성 정보를 통해 최종 텍스트 정보를 생성함으로써, 사용자의 발화 음성에 대한 인식률이 향상될 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 개시의 일 실시 예에 따른, 사용자 음성에 응답하여 동작을 수행하는 전자 장치의 사용도이다.

도 2는 본 개시의 일 실시 예에 따른, 전자 장치의 구성을 간략히 도시한 블록도 이다.

도 3은 본 개시의 일 실시 예에 따른, 전자 장치의 구성을 상세히 도시한 블록도이다.

도 4은 본 개시의 일 실시 예에 따른, 인공지능 에이전트 시스템의 대화 시스템을 도시한 블록도이다.

도 5는 본 개시의 일 실시 예에 따른, 전자 장치의 제어 방법을 설명하기 위한 흐름도 이다.

도 6은 본 개시의 일 실시 예에 따른, 사용자의 발화에 대응되는 최종 텍스트 정보를 획득하는 방법을 나타내는 도면이다.

도 7a는 본 개시의 일 실시 예에 따른, 사용자 발화 음성에 대응되는 최종 텍스트 정보를 바탕으로, 동작을 수행할 장치를 식별하는 방법을 설명하기 위한 도면이다.

도 7b는 본 개시의 일 실시 예에 따른, 사용자 발화 음성에 대응되는 최종 텍스트 정보를 바탕으로, 동작을 수행할 장치를 식별하는 방법을 설명하기 위한 도면이다.

도 7c는 본 개시의 일 실시 예에 따른, 사용자 발화 음성에 대응되는 최종 텍스트 정보를 바탕으로, 동작을 수행할 장치를 식별하는 방법을 설명하기 위한 도면이다.

도 8은 본 개시의 일 실시 예에 따른 전자 장치 및 복수의 외부 장치간의 동작을 설명하기 위한 시퀀스도이다.

도 9은 본 개시의 일 실시 예에 따른 전자 장치 및 복수의 외부 장치간의 동작을 설명하기 위한 시퀀스도이다.

발명을 실시하기 위한 구체적인 내용

[0012] 이하에서는 도면을 참조하여, 본 개시에 대해 상세히 설명하기로 한다. 도 1은 본 개시의 일 실시 예에 따른, 사용자 음성에 응답하여 동작을 수행하는 전자 장치의 사용도이다.

[0013] 전자 장치(100)는 인공 지능 에이전트 프로그램을 활성화하기 위한 트리거 음성을 입력받을 수 있다. 예로, 트리거 음성인 “빅스비”, “하이 빅스비” 등과 같은 트리거 단어를 포함할 수 있으며, 도 1에는 트리거 음성이 “AAA”인 경우를 도시하였다. 트리거 음성에 응답하여 전자 장치(100)는 인공지능 에이전트 프로그램을 활성화시킬 수 있다. 인공지능 에이전트 프로그램은 사용자 음성에 대한 응답을 자연어로 처리하여 제공하며, 사용자 음성에 대한 동작을 수행할 수 있는 대화 시스템을 포함할 수 있다. 또한, 일 실시 예로, 전자 장치(100)는 인공 지능 에이전트 프로그램을 활성화하기 위한 트리거 단어 이외에 전자 장치(100)에 구비된 특정 버튼을 선택한 후, 사용자 음성을 입력받을 수 있다. 또한, 일 실시 예로, 태내에 위치한 적어도 하나의 외부 장치가 트리거 음성을 수신한 경우, 전자 장치(100)가 트리거 음성을 수신한 외부 장치로부터 음성 인식 모드로 진입하기 위한 정보를 수신하여, 음성 인식 모드로 진입할 수 있다. 본 개시의 일 실시 예에 따른, 적어도 하나의 외부 장치와 전자 장치(100)는 인공지능 에이전트가 동일할 수 있다. 즉, 태내에 위치한 적어도 하나의 외부 장치와 전자 장치(100)는 동일한 트리거 음성을 인식하여, 음성 인식 모드로 진입할 수 있다.

[0014] 트리거 음성이 수신되면, 전자 장치(100)는 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 인공지능 에이전트 프로그램을 활성화시킬 수 있다. 본 개시의 일 실시 예로, 트리거 음성이 수신되면, 전자 장치(100)는 음성 인식 모드로 진입하기 위한 정보를 태내에 위치한 적어도 하나의 외부 장치로 전송할 수 있다. 태내에 위치한 적어도 하나의 외부 장치는 전자 장치(100)로부터 음성 인식 모드로 진입하기 위한 정보를 수신하여, 음성 인식 모드로 진입할 수 있다. 상술한 과정을 통해, 태내에 위치한 적어도 하나의 외부 장치가 트리거 음성을 수신한 경우, 태내의 인공지능 에이전트 프로그램을 포함하고 있는 모든 외부 장치가 음성 인식 모드로 진입할 수 있게 된다.

[0015] 인공지능 에이전트 프로그램이 활성화되어 전자 장치(100)가 음성 인식모드로 동작하는 동안, 전자 장치(100)는 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득할 수 있다. 이때, 사용자가 발화한 음성은 특정 동작을 수행하기 위한 음성일 수 있다. 예로, 도 1에 도시된 바와 같이, 전자 장치(100)는 “에어컨 온도 25도로 낮춰줘” 라는 사용자 음성에 대응되는 제1 음성 정보를 획득할 수 있다.

[0016] 그리고, 전자 장치(100)는 태내에 위치한 적어도 하나의 외부 장치로부터 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득할 수 있다. 예로, 도 1을 참조하면, 사용자가 “AAA” 라는 트리거 음성을 발화한 후, “에어컨 온도 25도로 낮춰줘” 라고 발화하면, 제1 외부 장치(10)는 사용자 발화에 대응되는 제2-1 음성 정보를 획득하고, 제2 외부 장치(20)는 사용자 발화에 대응되는 제2-2 음성 정보를 획득할 수 있다. 또한, 제3 외부 장치(30)는 사용자 발화에 대응되는 제2-3 음성 정보를 획득할 수 있다. 즉, 동일한 사용자의 발화 음성에 대해 외부 장치 각각이 사용자의 발화한 음성에 대응되는 신호를 수신하고, 외부 장치 각각은 수신한 신호

에 대응되는 음성 정보를 전자 장치(100)로 전송할 수 있다.

- [0017] 일 실시 예에 따르면, 제2-1 음성 정보 내지 제2-3 음성 정보는 서로 상이할 수 있다. 즉, 적어도 하나의 외부 장치와 사용자의 거리, 외부 환경 등에 따라, 동일한 사용자의 발화에 대해서도 외부 장치 각각이 수신한 음성 정보가 각기 상이할 수 있다. 예로, “에어컨 온도 25도로 낮춰줘” 라는 사용자 음성에 대해 제1 외부 장치(10)는 제1 외부 장치(10)에서 출력되는 오디오에 의해, “에어컨 온도 15도로 낮춰줘” 에 대응되는 제2-1 음성 정보를 획득할 수 있다. 그리고, 제2 외부 장치(20)는 사용자와 멀리 떨어져 있어, “에어컨 15도” 에 대응되는 제2-2 음성 정보를 획득할 수 있다.
- [0018] 전자 장치(100)에 포함된 마이크로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득하고, 적어도 하나의 외부 장치로부터 적어도 하나의 제2 음성 정보를 획득하면, 전자 장치(100)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다. 구체적으로, 전자 장치(100)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 ASR 모듈에 입력하여, 제1 음성 정보에 대응되는 제1 텍스트 정보 및 적어도 하나의 제2 음성 정보에 대응되는 적어도 하나의 제2 텍스트 정보를 획득할 수 있다. 그리고, 전자 장치(100)는 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 획득할 수 있다. 도 1을 참조하면, 전자 장치(100)는 전자 장치에 포함된 마이크로부터 획득한 제1 음성 정보, 제1 외부 장치로부터 획득한 제2-1 음성 정보, 제2 외부 장치로부터 획득한 제2-2 음성 정보, 및 제3 외부 장치로부터 획득한 제2-3 음성 정보를 ASR 모듈에 입력하여 각각의 음성 정보에 대응되는 텍스트 정보를 획득하고, 획득된 복수의 텍스트 정보를 통해, “에어컨 온도 25도로 낮춰줘” 라는 최종 텍스트 정보를 획득할 수 있다. 다만, 이에 한정되지 않고, 전자 장치(100)는 적어도 하나의 외부 장치로부터 해당 외부 장치가 획득한 음성 정보에 대응되는 적어도 하나의 텍스트 정보를 수신하고, 수신한 복수의 텍스트 정보 및 제1 음성 정보에 대응되는 텍스트 정보를 통해, 최종 텍스트 정보를 획득할 수 있다.
- [0019] 그리고, 전자 장치(100)는 최종 텍스트 정보를 NLU 모듈에 입력하여, 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 획득할 수 있다. 또한, 최종 텍스트 정보는 적어도 하나의 키워드를 포함하고, 전자 장치(100)는 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 바탕으로, 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다.
- [0020] 다만, 이에 한정되지 않으며, 최종 텍스트 정보는 사용자가 발화한 음성의 음절 각각에 대응되는 복수의 키워드를 포함할 수 있다. 그리고, 전자 장치(100)는 복수의 키워드를 바탕으로 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다. 즉, 일 실시 예로, 전자 장치(100)는 사용자가 발화한 음성에 대한 모든 음절을 이용하여 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다.
- [0021] 즉, 도 1을 참조하면, “에어컨 온도 25도로 낮춰줘” 라는 최종 텍스트 정보를 통해, 전자 장치(100)는 에어컨의 설정 온도를 25도로 설정하는 동작에 대한 정보를 획득할 수 있고, 이러한 동작을 수행할 외부 장치를 공기정화 장치인 제2 외부 장치(20)로 식별할 수 있다. 다만, 이에 한정되지 않으며, 전자 장치(100)는 사용자가 발화한 음성에 대응되는 동작을 수행할 장치가 전자 장치(100)인 것으로도 식별할 수 있다. 사용자가 발화한 음성에 대응되는 동작에 대한 정보 및 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 방법에 대한 구체적인 내용은 도 7a 내지 도 7c 를 통해 상세히 설명하도록 한다.
- [0022] 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 외부 장치가 식별된 경우, 전자 장치(100)는 식별된 외부 장치로 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 전송할 수 있다. 그리고, 식별된 외부 장치는 사용자가 발화한 음성에 대응되는 동작을 수행할 수 있다. 즉, 도 1을 참조하면, 전자 장치(100)는 제2 외부 장치(20)로 설정 온도를 25도로 설정하기 위한 정보를 전송하고, 제2 외부 장치(20)는 설정 온도를 25도로 설정할 수 있다.
- [0023] 상술한 바와 같은 실시 예에 의해, 트리거 음성이 수신되면, 전자 장치(100)는 태크의 인공지능 에이전트 프로그램 포함하고 있는 모든 외부 장치를 음성 인식 모드로 진입하도록 제어함으로써, 사용자가 의도하지 않은 외부 장치만 음성 인식 모드로 진입하게 되는 문제점이 방지될 수 있다.
- [0024] 또한, 적어도 하나의 외부 장치로부터 수신한 음성 정보를 통해 최종 텍스트 정보를 생성함으로써, 사용자의 발화 음성에 대한 인식률이 향상될 수 있다.
- [0025] 한편, 상술한 실시 예에서는 전자 장치(100) 내에 인공지능 에이전트 시스템이 저장되어 전자 장치(100)가 직접 사용자 음성이 사용자 확인이 필요한 태스크인지 여부를 판단하고, 질문을 생성하는 것으로 설명하였으나, 이는

일 실시 예에 불과할 뿐, 상술한 동작 중 일부는 외부 서버에 의해 구현될 수 있다. 예로, 외부 서버는 사용자 음성에 대한 텍스트를 획득하거나 사용자 음성이 사용자 확인이 필요한 태스크인지 여부를 판단하거나 질문을 생성할 수 있다.

[0026] 한편, 전자 장치(100)는 상술한 바와 같은 사용자 음성에 대한 응답을 제공하기 위하여 인공지능 에이전트(Artificial intelligence agent)를 이용할 수 있다. 이때, 인공지능 에이전트는 AI(Artificial Intelligence) 기반의 서비스(예를 들어, 음성 인식 서비스, 비서 서비스, 번역 서비스, 검색 서비스 등)를 제공하기 위한 전용 프로그램으로서, 기존의 범용 프로세서(예를 들어, CPU) 또는 별도의 AI 전용 프로세서(예를 들어, GPU 등)에 의해 실행될 수 있다. 특히, 인공지능 에이전트는 후술할 다양한 모듈(예로, 대화 시스템)을 제어할 수 있다.

[0027] 구체적으로, 기설정된 사용자가 발화한 음성(예를 들어, "빅스비" 등)이 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러진 경우, 인공지능 에이전트가 동작할 수 있다. 그리고, 인공지능 에이전트는 사용자가 발화한 음성을 분석하여 음성에 대응되는 동작 및 동작을 수행할 주체를 식별할 수 있다.

[0028] 물론, 기설정된 트리거 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지면 인공지능 에이전트가 동작할 수도 있다. 또한, 인공지능 에이전트는 기설정된 트리거 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지기 이전에 기 실행된 상태일 수 있다. 이 경우, 기설정된 트리거 음성(예를 들어, "빅스비" 등)이 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러진 이후에는 전자 장치(100)의 인공지능 에이전트가 사용자 음성에 대한 태스크를 수행할 수 있다. 예를 들어, 인공지능 에이전트가 AI 전용 프로세서에 의해 실행되는 경우, 기설정된 사용자 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지기 전에는 범용 프로세서에 의해 전자 장치(100)의 기능이 실행되며, 기설정된 사용자 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러진 이후에는 AI 전용 프로세서에 의해 전자 장치(100)의 기능이 실행될 수 있다.

[0029] 또한, 인공지능 에이전트는 기설정된 사용자 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지기 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 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지면, 전자 장치(100)는 인공지능 에이전트를 동작시키고, 동작된 인공지능 에이전트를 이용하여 사용자 음성에 대한 태스크를 수행할 수 있다.

[0030] 또한, 인공지능 에이전트는 기설정된 사용자 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지기 이전에 종료된 상태일 수 있다. 인공지능 에이전트가 종료된 상태에서 기설정된 사용자 음성(예를 들어, "빅스비" 등)가 입력되거나 전자 장치(100)에 구비된 버튼(예를 들어, 인공지능 에이전트를 실행하기 위한 버튼)이 눌러지면, 전자 장치(100)는 인공지능 에이전트를 실행시키고, 실행된 인공지능 에이전트를 이용하여 사용자 음성에 대한 태스크를 수행할 수 있다.

[0031] 한편, 인공지능 에이전트는 후술할 다양한 장치 또는 모듈을 제어할 수 있다. 이에 대해서는 추후 상세히 설명하기로 한다.

[0032] 또한, 전자 장치(100) 및 서버 간의 학습된 다양한 모델을 이용하여 사용자 음성을 분석하여 사용자 확인이 필요한 태스크와 관련된 사용자 음성인지 여부를 판단하고 판단 결과를 바탕으로 질문을 생성하여 제공하는 구체적인 예들은 이하 다양한 실시예들을 통하여 후술된다.

[0034] 도 2는 본 개시의 일 실시 예에 따른, 전자 장치의 구성을 간략히 도시한 블록도 이다. 도 2에 도시된 바와 같이, 전자 장치(100)는 입력 인터페이스(110), 통신 인터페이스(120), 메모리(130) 및 프로세서(140)를 포함할 수 있다. 그러나, 상술한 구성에 한정되는 것은 아니며, 전자 장치의 유형에 따라 일부 구성이 추가되거나 생략

될 수 있음은 물론이다.

- [0035] 입력 인터페이스(110)는 전자 장치(100)를 제어하기 위한 사용자 입력을 수신할 수 있다. 예로, 입력 인터페이스(110)는 전자 장치(100)를 제어하기 위한 사용자의 사용자 터치, 사용자 발화 음성 등과 같은 다양한 사용자 조작을 입력받을 수 있다. 특히, 입력 인터페이스(110)는 태스크를 수행하기 위한 사용자 발화 음성을 입력받을 수 있다. 일 실시 예로, 입력 인터페이스(110)는 마이크를 포함할 수 있으며, 전자 장치(100)는 마이크를 통해 사용자 발화 음성을 입력 받을 수 있다.
- [0036] 통신 인터페이스(120)는 외부 전자 장치와 통신을 수행할 수 있다. 한편, 통신 인터페이스(120)가 외부 장치와 통신 연결되는 것은 제3 기기(예로, 중계기, 허브, 액세스 포인트, 서버 또는 게이트웨이 등)를 거쳐서 통신하는 것을 포함할 수 있다. 무선 통신은, 예를 들면, LTE, LTE-A(LTE Advance), CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), 또는 GSM(Global System for Mobile Communications) 등 중 적어도 하나를 사용하는 셀룰러 통신을 포함할 수 있다. 일 실시예에 따르면, 무선 통신은, 예를 들면, WiFi(wireless fidelity), 블루투스, 블루투스 저전력(BLE), 지그비(Zigbee), NFC(near field communication), 자력 시큐어 트랜스미션(Magnetic Secure Transmission), 라디오 프리퀀시(RF), 또는 보디 에어리어 네트워크(BAN) 중 적어도 하나를 포함할 수 있다. 유선 통신은, 예를 들면, USB(universal serial bus), HDMI(high definition multimedia interface), RS-232(recommended standard232), 전력선 통신, 또는 POTS(plain old telephone service) 등 중 적어도 하나를 포함할 수 있다. 무선 통신 또는 유선 통신이 수행되는 네트워크는 텔레커뮤니케이션 네트워크, 예를 들면, 컴퓨터 네트워크(예: LAN 또는 WAN), 인터넷, 또는 텔레폰 네트워크 중 적어도 하나를 포함할 수 있다.
- [0037] 또한, 통신 인터페이스(120)는 외부의 서버와 통신을 수행하여 인공지능 에이전트 서비스를 제공할 수 있다. 특히, 통신 인터페이스(120)는 외부 서버로 사용자 발화 음성에 대응되는 음성 정보를 전송할 수 있으며, 외부 서버로부터 사용자 발화 음성에 대응되는 동작 및 동작을 수행할 장치에 대한 정보를 수신할 수 있다.
- [0038] 메모리(130)는 전자 장치(100)의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 특히, 메모리(130)는 비휘발성 메모리, 휘발성 메모리, 플래시메모리(Flash-memory), 하드디스크 드라이브(HDD) 또는 솔리드 스테이트 드라이브(SSD) 등으로 구현될 수 있다. 메모리(130)는 프로세서(140)에 의해 액세스되며, 프로세서(140)에 의한 데이터의 독취/기록/수정/삭제/갱신 등이 수행될 수 있다. 본 개시에서 메모리라는 용어는 메모리(130), 프로세서(140) 내 롬(미도시), 램(미도시) 또는 전자 장치(100)에 장착되는 메모리 카드(미도시)(예를 들어, micro SD 카드, 메모리 스틱)를 포함할 수 있다. 또한, 메모리(130)에는 디스플레이의 디스플레이 영역에 표시될 각종 화면을 구성하기 위한 프로그램 및 데이터 등이 저장될 수 있다.
- [0039] 또한, 메모리(130)는 대화 시스템을 동작하기 위한 인공지능 에이전트를 저장할 수 있다. 구체적으로, 전자 장치(100)는 사용자 발화에 대한 응답으로 자연어를 생성하기 위하여 인공지능 에이전트(Artificial intelligence agent)를 이용할 수 있다. 이때, 인공지능 에이전트는 AI(Artificial Intelligence) 기반의 서비스(예를 들어, 음성 인식 서비스, 비서 서비스, 번역 서비스, 검색 서비스 등)를 제공하기 위한 전용 프로그램이다. 특히, 인공지능 에이전트는 기존의 범용 프로세서(예를 들어, CPU) 또는 별도의 AI 전용 프로세서(예를 들어, GPU 등)에 의해 실행될 수 있다.
- [0040] 또한, 메모리(130)는 도 4에 도시된 바와 같은 대화 시스템을 구성하는 복수의 구성(또는 모듈)을 포함할 수 있다. 이에 대해서는 도 4를 참조하여 상세히 설명하기로 한다.
- [0041] 프로세서(140)는 메모리(130)와 전기적으로 연결되어 전자 장치(100)의 전반적인 동작 및 기능을 제어할 수 있다. 특히, 프로세서(140)는 메모리(130)에 저장된 적어도 하나의 명령을 실행함으로써, 입력 인터페이스(110)를 통해, 사용자의 트리거 음성을 수신할 수 있다. 다만, 이에 한정되지 않고, 통신 인터페이스(120)를 통해 택내에 위치한 외부 장치로부터 음성 인식 모드로 진입하기 위한 정보를 수신할 수 있다.
- [0042] 트리거 음성 또는 음성 인식 모드로 진입하기 위한 정보가 수신되면, 프로세서(140)는 음성 명령을 인식 하기 위한 음성 인식 모드로 진입하여, 인공 지능 에이전트를 활성화 시킬 수 있다. 그리고, 프로세서(140)는 음성 인식 모드로 진입하기 위한 정보를 택내에 위치한 적어도 하나의 외부 장치로 전송할 수 있다. 본 개시의 일 실시 예에 따르면, 프로세서(140)는 택내에 위치한 적어도 하나의 외부 장치로 음성 인식 모드로 진입하기 위한 정보와 함께 외부 장치를 음소거 모드로 진입하기 위한 정보를 전송할 수 있다. 일 실시 예로, 프로세서(140)는 적어도 하나의 외부 장치 중 영상 또는 오디오가 재생 중인 외부 장치로 음성 인식 모드로 진입하기 위한 정보와 함께 외부 장치를 음소거 모드로 진입하기 위한 정보를 전송할 수 있다. 또한, 일 실시 예로, 전자 장치

(100) 및 적어도 하나의 외부 장치에 포함된 음성 인식 엔진은 동일할 수 있다. 즉, 전자 장치(100) 및 적어도 하나의 외부 장치는 동일한 인공지능 에이전트 프로그램을 포함할 수 있다.

[0043] 음성 인식 모드로 동작하는 동안, 프로세서(140)는 입력 인터페이스(110)에 포함된 마이크로부터 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득하고, 적어도 하나의 외부 장치로부터 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득할 수 있다. 일 실시 예에 따르면, 제1 음성 정보 및 적어도 하나의 제2 음성 정보는 동일한 사용자의 발화에 대응되는 음성 정보일 수 있으며, 제1 음성 정보는 전자 장치(100)가 마이크를 통해 수신한 음성 정보며, 적어도 하나의 제2 음성 정보는 적어도 하나의 외부 장치가 해당 외부 장치의 마이크를 통해 수신할 음성 정보일 수 있다. 따라서, 사용자와 외부 장치의 위치, 환경에 따라 제1 음성 정보 및 적어도 하나의 제2 음성 정보는 상이할 수 있다.

[0044] 그리고, 프로세서(140)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 장치를 식별할 수 있다. 구체적으로, 프로세서(140)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 자동 음성 입력 모듈(ASR) 모듈에 입력하여, 제1 음성 정보에 대응되는 제1 텍스트 정보 및 적어도 하나의 제2 음성 정보에 대응되는 적어도 하나의 제2 텍스트 정보를 획득할 수 있다. 그리고, 프로세서(140)는 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 획득할 수 있다.

[0045] 본 개시의 일 실시 예로, 제1 텍스트 정보와 적어도 하나의 제2 텍스트 정보가 동일한 경우, 프로세서(140)는 제1 텍스트 정보를 최종 텍스트 정보로 식별할 수 있다. 제1 텍스트 정보와 적어도 하나의 제2 텍스트 정보 중 적어도 두 개가 서로 상이한 경우, 프로세서(140)는 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 인공 신경망 모델에 입력하여, 최종 텍스트 정보를 획득할 수 있다. 본 개시에 따른 인공 신경망 모델은 복수의 텍스트 정보를 입력 받아, 사용자의 발화에 대응되는 최종 텍스트 정보를 생성하는 신경망 모델일 수 있다. 본 개시에 따른 인공 신경망 모델은, 복수의 신경망 레이어들로 구성될 수 있다. 각 레이어는 복수의 가중치(weight values)을 갖고 있으며, 이전(previous) 레이어의 연산 결과와 복수의 가중치의 연산을 통해 레이어의 연산을 수행한다. 신경망의 예로는, CNN (Convolutional Neural Network), DNN (Deep Neural Network), RNN (Recurrent Neural Network), RBM (Restricted Boltzmann Machine), DBN (Deep Belief Network), BRDNN(Bidirectional Recurrent Deep Neural Network) 및 심층 Q-네트워크(Deep Q-Networks)이 있으며, 본 개시에서의 신경망은 명시한 경우를 제외하고 전술한 예에 한정되지 않는다.

[0046] 일 실시 예로, 본 개시에 따른 인공 신경망 모델은 사용자의 명령어 내역, 입력된 복수의 텍스트 정보에 포함된 중복된 단어 및 복수의 텍스트 정보간의 유사도를 바탕으로, 복수의 텍스트 정보에 대응되는 최종 텍스트 정보를 획득할 수 있다.

[0047] 일 실시 예로, 프로세서(140)는 사용자가 발화한 명령어 목록을 획득할 수 있다. 사용자가 발화한 명령어 목록에는 사용자가 자주 발화한 상위 100개의 명령어에 대한 정보가 포함될 수 있다. 그리고, 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보가 상이한 경우, 인공 신경망 모델은 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 명령어 목록에 포함된 복수의 명령어 각각과 유사도를 식별할 수 있다. 식별된 유사도를 바탕으로, 프로세서(140)는 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보 중 유사도가 가장 높은 텍스트 정보를 최종 텍스트 정보로 식별할 수 있다. 그리고, 프로세서(140)는 최종 텍스트 정보를 자연어 이해 모듈(NLU) 모듈에 입력하여, 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 획득할 수 있다. 본 개시에 따른 일 실시 예로, 구체적으로, 프로세서(140)는 최종 텍스트 정보를 자연어 이해 모듈(NLU)에 입력하여 사용자가 발화한 음성에 대응되는 최종 텍스트 정보에 대한 사용자 의도 및 슬롯에 대한 정보를 획득할 수 있다. 자연어 이해 모듈(NLU)에 대한 구체적인 내용은 도 4를 통해 후술하도록 한다.

[0048] 또한, 최종 텍스트 정보는 적어도 하나의 키워드를 포함하고, 프로세서(140)는 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 바탕으로, 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다. 일 실시 예로, 프로세서(140)는 최종 텍스트 정보에 포함된 적어도 하나의 키워드 각각에 부여된 가중치를 통해 동작을 수행할 외부 장치를 식별할 수 있다.

[0049] 또한, 최종 텍스트 정보는 상기 사용자가 발화한 음성의 음절 각각에 대응되는 복수의 키워드를 포함하고, 프로세서(140)는 최종 텍스트 정보에 포함된 복수의 키워드를 바탕으로, 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다. 사용자가 발화한 음성에 대응되는 동작에 대한 정보 및 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별하는 방법에 대한 구체적인 내용은 도 7a 내지 도 7c 를 통해 상세히 설명하도록 한다.

- [0050] 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 외부 장치가 식별된 경우, 프로세서(140)는 식별된 외부 장치로 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 전송할 수 있다. 그리고, 식별된 외부 장치는 사용자가 발화한 음성에 대응되는 동작을 수행할 수 있다.
- [0052] 도 3은 본 개시의 일 실시 예에 따른, 전자 장치의 구성을 상세히 도시한 블록도이다. 도 3에 도시된 바와 같이, 전자 장치(100)는 입력 인터페이스(110), 통신 인터페이스(120), 메모리(130), 디스플레이(150), 스피커(160), 센서(170) 및 프로세서(140)를 포함할 수 있다. 한편, 도 3에 도시된 입력 인터페이스(110), 통신 인터페이스(120), 메모리(130) 및 프로세서(140)는 도 2에서 설명하였으므로, 중복되는 설명은 생략하기로 한다.
- [0053] 입력 인터페이스(110)는 전자 장치(100)를 제어하기 위한 사용자 입력을 수신할 수 있다. 특히, 입력 인터페이스(110)는 특정 동작을 수행하기 위한 사용자 음성을 입력받을 수 있다. 입력 인터페이스(110)는 도 3에 도시된 바와 같이, 사용자 음성을 입력받기 위한 마이크(111), 사용자 손 또는 스타일러스 펜 등을 이용한 사용자 터치를 입력받기 위한 터치 패널(113), 사용자 조작을 입력받기 위한 버튼(115) 등이 포함될 수 있다. 그러나, 도 3에 도시된 입력 인터페이스(110)의 예는 일 실시예에 불과할 뿐, 다른 입력 장치(예로, 키보드, 마우스, 모션 입력부 등)로 구현될 수 있다.
- [0054] 디스플레이(150)는 프로세서(140)의 제어에 따라 다양한 정보를 표시할 수 있다. 특히, 디스플레이(150)는 복수의 오브젝트를 포함하는 UI를 표시할 수 있다. 또한, 디스플레이(150)는 사용자와 인공지능 에이전트 시스템 간의 대화를 포함하는 메시지 창을 표시할 수 있다. 디스플레이(150)는 터치 패널(113)과 함께 터치 스크린으로 구현될 수 있다.
- [0055] 스피커(160)는 오디오 처리부에 의해 디코딩이나 증폭, 노이즈 필터링과 같은 다양한 처리 작업이 수행된 각종 오디오 데이터뿐만 아니라 각종 알람 음이나 음성 메시지를 출력하는 구성이다. 특히, 스피커(160)는 사용자가 발화한 음성에 대한 동작 또는 안내 메시지를 자연어 형태의 음성 메시지로 출력할 수 있다. 한편, 오디오를 출력하기 위한 구성은 스피커로 구현될 수 있으나, 이는 일 실시 예에 불과할 뿐, 오디오 데이터를 출력할 수 있는 출력 단자로 구현될 수 있다.
- [0056] 센서(170)는 전자 장치(100)의 다양한 상태 정보를 감지할 수 있다. 예로, 센서(170)는 전자 장치(100)의 움직임 정보를 감지할 수 있는 움직임 센서(예로, 자이로 센서, 가속도 센서 등)를 포함할 수 있으며, 위치 정보를 감지할 수 있는 센서(예로, GPS(Global Positioning System) 센서), 전자 장치(100) 주위의 환경 정보를 감지할 수 있는 센서(예로, 온도 센서, 습도 센서, 기압 센서 등), 전자 장치(100)의 사용자 정보를 감지할 수 있는 센서(예로, 혈압 센서, 혈당 센서, 맥박수 센서 등) 등을 포함할 수 있다. 그 밖에, 센서(170)는 전자 장치(100)의 외부를 촬영하기 위한 이미지 센서 등을 더 포함할 수 있다.
- [0058] 도 4은 본 개시의 일 실시 예에 따른, 인공지능 에이전트 시스템의 대화 시스템을 도시한 블록도이다. 도 4에 도시된 대화 시스템(400)은 가상의 인공지능 에이전트와 자연어를 통해 대화를 수행하거나 전자 장치(100)를 제어하기 위한 구성으로서, 본 개시의 일 실시예 따르면, 대화 시스템(400)에 포함된 모듈은 전자 장치(100)의 메모리(130) 내에 저장될 수 있으나, 이는 일 실시 예에 불과할 뿐, 하드웨어와 소프트웨어의 결합된 형태로 구현될 수 있다. 또한, 대화 시스템(400)에 포함된 적어도 하나의 모듈은 외부의 적어도 하나의 서버에 포함될 수 있다.
- [0059] 대화 시스템(200)은 도 4에 도시된 바와 같이, 음성 입력(speech input) 모듈(410), 자동 음성 인식(Auto speech recognition)(ASR) 모듈(420), 자연어 이해(natural language understanding)(NLU) 모듈(430), 대화 매니저(dialogue manager)(DM) 모듈(440), 액션 플래너(action planner)(AP) 모듈(450), 자연어 생성(natural language generator)(NLG) 모듈(460), 텍스트 음성 변환(text to speech)(TTS) 모듈(470), 출력(output) 모듈(480) 및 기기 제어(device control) 모듈(490)을 포함할 수 있다.
- [0060] 음성 입력 모듈(410)은 음성 데이터 형태의 사용자 발화 음성을 입력 받을 수 있다. 이때, 음성 입력 모듈(410)은 마이크를 포함할 수 있으며, 마이크를 통해 사용자 발화 음성을 포함하는 아날로그 형태의 오디오 신호를 수신할 수 있으며, 아날로그 신호를 디지털 신호로 변환할 수 있다. 일 실시 예로, 음성 입력 모듈(410)은 디지털 신호로부터 노이즈 성분을 제거하여 사용자 발화 음성에 대응되는 음성 정보를 획득할 수 있다. 다만, 다른 실시 예에 따르면, 음성 입력 모듈(410)이 별도의 노이즈 성분을 제거하지 않고, 이후 음성 인식 모델을 통해

디지털 신호에 포함된 노이즈가 제거될 수 있다. 특히, 음성 입력 모듈(410)을 통해 수신되는 사용자 발화 음성은 적어도 하나의 텍스트를 포함하는 문장 또는 구 형태일 수 있으나, 이는 일 실시 예에 불과할 뿐, 복수의 문장 또는 구를 포함할 수 있다.

[0061] 한편, 음성 입력 모듈(410)에 포함된 마이크는 전자 장치(100) 내부에 구비될 수 있으나, 이는 일 실시 예에 불과할 뿐, 외부에 구비되어 전자 장치(100)와 유선 또는 무선으로 연결될 수 있다. 또한, 음성 입력 모듈(410)은 외부 장치로부터 사용자 음성을 수신하는 통신 모듈을 포함할 수 있음은 물론이다.

[0062] 자동 음성 인식 모듈(Auto speech recognition)(ASR)(420)은 음성 입력 모듈(410)로부터 수신된 음성 정보 형태의 사용자 발화 음성을 텍스트 정보로 변환할 수 있다. 이때, 텍스트 정보는 문자 코드로 이루어진 데이터이며, 음성 정보는 사용자 발화 음성에 대한 주파수 정보를 포함하는 데이터일 수 있다. 자동 음성 인식 모듈(420)은 음향(acoustic) 모델 및 언어(language) 모델을 포함할 수 있다. 이때, 음향 모델은 발성에 관련된 정보를 포함할 수 있고, 언어 모델은 단위 음소 정보 및 단위 음소 정보의 조합에 대한 정보를 포함할 수 있다. 음성 인식 모듈은 발성에 관련된 정보 및 단위 음소 정보에 대한 정보를 이용하여 사용자 발화 음성을 텍스트 정보로 변환할 수 있다. 음향 모델 및 언어 모델에 대한 정보는, 예를 들어, 자동 음성 인식 데이터베이스(automatic speech recognition database)(ASR DB)에 저장될 수 있다. 본 개시에 따른 일 실시 예로, 자동 음성 인식 모듈(420)은 제1 음성 정보에 대응되는 제1 텍스트 정보, 적어도 하나의 제2 음성 정보에 대응되는 제2 텍스트 정보를 생성할 수 있다. 그리고, 자동 음성 인식 모듈(420)은 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 입력 받아, 최종 텍스트 정보를 생성하기 위한 인공지능 신경망 모델을 포함할 수 있다.

[0063] 자연어 이해 모듈(natural language understanding)(NLU)(230)은 자동 음성 인식 모듈(220)에서 변환된 텍스트 정보에 대해 문법적 분석(syntactic analyze) 또는 의미적 분석(semantic analyze)을 수행하여 사용자 발화 음성에 대한 도메인 및 사용자 의도를 파악할 수 있다. 이때, 문법적 분석은 사용자 입력을 문법적 단위(예: 단어, 구, 형태소 등)로 나누고, 나누어진 단위가 어떤 문법적인 요소를 갖는지 파악할 수 있다. 의미적 분석은 의미(semantic) 매칭, 룰(rule) 매칭, 포물러(formula) 매칭 등을 이용하여 수행할 수 있다.

[0064] 본 개시에 따른 일 실시 예로, 최종 텍스트 정보가 자연어 이해 모듈(NLU)(40)에 입력되어, 사용자가 발화한 음성에 대응되는 동작에 대한 정보가 획득될 수 있다. 구체적으로, 프로세서(140)는 최종 텍스트 정보를 자연어 이해 모듈(NLU)(40)에 입력하여 사용자가 발화한 음성에 대응되는 최종 텍스트 정보에 대한 사용자 의도 및 슬롯에 대한 정보를 획득할 수 있다. 즉, 본 개시에 따른 일 실시 예로, 사용자가 발화한 음성에 대응되는 동작에 대한 정보는 사용자의 의도 및 슬롯에 대한 정보를 포함할 수 있다.

[0065] 대화 매니저 모듈(440)은 자연어 이해 모듈(430)을 통해 획득된 최종 텍스트 정보에 포함된 사용자 의도 및 슬롯에 대한 정보를 바탕으로 사용자 발화 음성에 대한 응답을 제공할 수 있다. 이때, 대화 매니저 모듈(440)은 지식 베이스를 기반으로 사용자 발화 음성에 대한 응답을 제공할 수 있다. 이때, 지식 베이스는 전자 장치(100) 내에 포함될 수 있으나, 이는 일 실시 예에 불과할 뿐, 외부 서버에 포함될 수 있다. 그리고, 대화 매니저 모듈(440)은 최종 텍스트 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 획득할 수 있다.

[0066] 자연어 생성 모듈(460)은 대화 매니저 모듈(440)에서 획득된 사용자 발화한 음성에 대응되는 동작에 대한 정보를 텍스트 형태로 변경할 수 있다. 텍스트 형태로 변경된 정보는 자연어의 형태일 수 있다. 사용자 발화한 음성에 대응되는 동작에 대한 정보는, 예를 들어, 추가 입력에 대한 정보, 사용자 입력에 대응되는 동작의 완료를 안내하는 정보, 질문에 대한 응답 정보 또는 사용자의 추가 입력을 안내하는 정보(예: 사용자 입력에 대한 피드백 정보)일 수 있다. 텍스트 형태로 변경된 정보는 전자 장치(100)의 디스플레이에 표시되거나, 텍스트 음성 변환 모듈(470)에 의해 음성 형태로 변경될 수 있다.

[0067] 텍스트 음성 변환 모듈(470)은 텍스트 형태의 정보를 음성 형태의 정보로 변경할 수 있다. 텍스트 음성 변환 모듈(470)은 자연어 생성 모듈(460)로부터 텍스트 데이터 형태의 정보를 수신하고, 텍스트 데이터 형태의 정보를 음성 데이터 형태의 정보로 변경할 수 있다.

[0068] 출력 모듈(480)은 텍스트 음성 변환 모듈(470)로부터 수신된 음성 데이터 형태의 정보를 출력할 수 있다. 이때, 출력 모듈(480)은 스피커 또는 음성 출력 단자를 통해 음성 데이터 형태의 정보를 출력할 수 있다. 또는 출력 모듈(480)은 자연어 생성 모듈(460)을 통해 획득된 텍스트 데이터 형태의 정보를 디스플레이 또는 영상 출력 단자를 통해 출력할 수 있다.

[0069] 액션 플래너 모듈(450)은 사용자 의도 및 슬롯에 대한 정보를 이용하여 적어도 하나의 액션 룰(또는 패스 룰(path rule))을 생성할 수 있다. 예를 들어, 액션 플래너 모듈(260)은 사용자 의도 및 슬롯에 대한 정보에 기초

하여 실행될 어플리케이션 및 어플리케이션에서 실행될 동작을 결정하여 적어도 하나의 액션 룰을 생성할 수 있다. 그리고, 액션 플래너 모듈(450) 실행될 어플리케이션 및 어플리케이션에서 실행될 동작을 바탕으로, 동작을 실행할 외부 장치를 식별할 수 있다. 일 실시 예로, 액션 플래너 모듈(450)은 최종 텍스트 정보에 포함된 사용자의 의도 및 슬롯을 바탕으로, 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 식별하고, 식별된 키워드 각각에 부여된 가중치를 통해, 동작을 수행할 외부 장치를 식별할 수 있다. 또한, 일 실시 예로, 액션 플래너 모듈(450)은 최종 텍스트 정보에 포함된 사용자의 의도 및 슬롯을 바탕으로, 최종 텍스트 정보에 포함된 사용자가 발화한 음성의 음절 각각에 대응되는 복수의 키워드를 식별하고, 식별된 복수의 키워드 각각에 부여된 가중치를 통해, 동작을 수행할 외부 장치를 식별할 수 있다. 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 바탕으로, 동작을 수행할 외부 장치가 결정되는 방법에 대해서는 도 7a 내지 도 7c를 통해 후술하도록 한다.

[0070] 기기 제어 모듈(490)은 액션 플래너 모듈(450)에 의해 생성된 적어도 하나의 액션 룰에 따라 전자 장치(100)를 제어할 수 있다. 예를 들어, 액션 플래너 모듈(450)에서 식별된 동작을 수행할 외부 장치가 전자 장치(100)인 경우, 기기 제어 모듈(490)은 전자 장치(100)가 식별된 동작에 따른 기능을 수행하도록 제어할 수 있다. 액션 플래너 모듈(450)에서 식별된 동작을 수행할 외부 장치가 맥내에 위치한 외부 장치인 경우, 기기 제어 모듈(490)은 외부 장치로 동작에 대한 정보를 외부 장치로 전송하도록 전자 장치(100)를 제어할 수 있다.

[0072] 도 5는 본 개시의 일 실시 예에 따른, 전자 장치의 제어 방법을 설명하기 위한 흐름도 이다.

[0073] 도 5를 참조하면, 전자 장치(100)는 사용자의 트리거 음성을 수신할 수 있다(S510). 일 실시 예로, 트리거 음성은 “빅스비”, “하이 빅스비” 등과 같은 트리거 단어를 포함할 수 있으며, 트리거 음성에 응답하여 전자 장치(100)는 인공지능 에이전트 프로그램을 활성화시킬 수 있다.

[0074] 트리거 음성이 수신되면, 전자 장치(100)는 음성 인식 모드로 진입하고, 음성 인식 모드로 진입하기 위한 정보를 적어도 하나의 외부 장치로 전송할 수 있다(S520). 구체적으로, 트리거 음성이 수신되면, 전자 장치(100)는 사용자의 음성 명령을 인식하기 위한 음성 인식 모드로 진입하고, 인공지능 에이전트 프로그램을 활성화 시킬 수 있다. 그리고, 트리거 음성이 수신되면, 전자 장치(100)는 음성 인식 모드로 진입하기 위한 정보를 맥내에 위치한 적어도 하나의 외부 장치로 전송할 수 있다. 맥내에 위치한 적어도 하나의 외부 장치는 전자 장치(100)로부터 음성 인식 모드로 진입하기 위한 정보를 수신하여, 음성 인식 모드로 진입할 수 있다. 이에 따라, 맥내에 위치한 적어도 하나의 외부 장치가 트리거 음성을 수신한 경우, 맥내의 모든 외부 장치가 음성 인식 모드로 진입할 수 있게 된다.

[0075] 인공지능 에이전트 프로그램이 활성화되어 전자 장치(100)가 음성 인식모드로 동작하는 동안, 전자 장치(100)는 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득할 수 있다(S530).

[0076] 그리고, 전자 장치(100)는 적어도 하나의 외부 장치로부터 사용자가 발화한 음성에 대응되는 적어도 하나의 제2 음성 정보를 획득할 수 있다(S540). 일 실시 예로, 적어도 하나의 외부 장치는 전자 장치(100)와 같은 맥내에 위치한 외부 장치일 수 있다.

[0077] 그리고, 전자 장치(100)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다(S550). 구체적으로, 전자 장치(100)는 제1 음성 정보 및 적어도 하나의 제2 음성 정보를 ASR 모듈에 입력하여, 제1 음성 정보에 대응되는 제1 텍스트 정보 및 적어도 하나의 제2 음성 정보에 대응되는 적어도 하나의 제2 텍스트 정보를 획득할 수 있다. 그리고, 전자 장치(100)는 제1 텍스트 정보 및 적어도 하나의 제2 텍스트 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 획득할 수 있다. 그리고, 전자 장치(100)는 최종 텍스트 정보를 NLU 모듈에 입력하여, 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 획득할 수 있다. 또한, 최종 텍스트 정보는 적어도 하나의 키워드를 포함하고, 전자 장치(100)는 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 바탕으로, 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다.

[0078] 사용자가 발화한 음성에 대응되는 동작 및 동작을 수행할 외부 장치가 식별된 경우, 전자 장치(100)는 식별된 외부 장치로 사용자가 발화한 음성에 대응되는 동작에 대한 정보를 전송할 수 있다(S560).

[0079] 상술한 바와 같은 실시 예에 의해, 트리거 음성이 수신되면, 전자 장치(100)는 맥내에 위치한 모든 외부 장치를 음성 인식 모드로 진입하도록 제어함으로써, 사용자가 의도하지 않은 외부 장치만 음성 인식 모드로 진입하게 되는 문제점이 방지될 수 있다.

- [0080] 또한, 적어도 하나의 외부 장치로부터 수신한 음성 정보를 통해 최종 텍스트 정보를 생성함으로써, 사용자의 발화 음성에 대한 인식률이 향상될 수 있다.
- [0082] 도 6은 본 개시의 일 실시 예에 따른, 사용자의 발화에 대응되는 최종 텍스트 정보를 획득하는 방법을 나타내는 도면이다.
- [0083] 도 6을 참조하면, 사용자는 “AAA” 라는 트리거 음성을 발화한 후, “오늘 미세먼지 어때?” 라는 음성을 발화할 수 있다. 그리고, 댁내에 위치한 전자 장치(100)는 사용자의 발화에 대응되는 제1 음성 정보를 획득하고, 제1 음성 정보를 통해 “오늘 미세먼지 어때?” 라는 제1 텍스트 정보를 획득할 수 있다. 그리고, 댁내에 위치한 제1 외부 장치(10)는 사용자의 발화에 대응되는 제2-1 음성 정보를 획득하고, 제2-1 음성 정보를 통해 “오늘 미세먼지 어때?” 라는 제2-1 텍스트 정보를 획득할 수 있다. 또한, 댁내에 위치한 제2 외부 장치(20)는 사용자의 발화에 대응되는 제2-2 음성 정보를 획득하고, 제2-2 음성 정보를 통해 “오늘 민서집 어때?” 라는 제2-2 텍스트 정보를 획득할 수 있다. 그리고, 댁내에 위치한 제3 외부 장치(30)는 사용자의 발화에 대응되는 제2-3 음성 정보를 획득하고, 제2-3 음성 정보를 통해 “오늘 미세먼지 어때?” 라는 제2-2 텍스트 정보를 획득할 수 있다. 도 6에서는 각각의 장치에서 텍스트 정보를 획득하는 것으로 도시하였으나, 이에 한정되지 않고, 각각의 장치가 수신한 음성 정보가 전자 장치(100)로 전송되어, 전자 장치(100)에서 수신한 복수의 음성 정보에 대응되는 텍스트 정보를 생성할 수 있다.
- [0084] 그리고, 전자 장치(100)는 복수의 외부 장치(10, 20, 30)로부터 제2-1 텍스트 정보 내지 제2-3 텍스트 정보를 수신하고, 제1 텍스트 정보 및 제2-1 텍스트 정보 내지 제2-3 텍스트 정보를 바탕으로, 사용자가 발화한 음성에 대응되는 최종 텍스트 정보를 “오늘 미세먼지 어때?” 로 식별할 수 있다.
- [0085] 상술한 실시 예를 통해, 복수의 외부 장치로부터 수신한 음성 정보를 통해 최종 텍스트 정보를 생성함으로써, 사용자의 발화 음성에 대한 인식률이 향상될 수 있다.
- [0087] 도 7a는 본 개시의 일 실시 예에 따른, 사용자 발화 음성에 대응되는 최종 텍스트 정보를 바탕으로, 동작을 수행할 장치를 식별하는 방법을 설명하기 위한 도면이다.
- [0088] 전자 장치(100)는 사용자 발화 음성에 대응되는 최종 텍스트 정보에 포함된 적어도 하나의 키워드를 통해, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치를 식별할 수 있다. 도 7a를 참조하면, 최종 텍스트 정보가 “아이유 노래 들어줘” 인 경우, 전자 장치(100)는 최종 텍스트 정보로부터 “아이유”, “노래”, “들어줘” 의 3개의 키워드를 식별할 수 있다. 그리고, 각각의 키워드와 관련된 외부 장치를 식별할 수 있다. 즉, 도 7a를 참조하면, “아이유” 키워드는 가수에 대한 키워드로 냉장고와는 관련도가 없는 것(예로, X)으로 식별되며, TV 및 스피커와는 관련도가 높은 것(예로, 0)으로 식별될 수 있다. 그리고, “노래” 는 오디오에 관한 키워드로 냉장고와는 관련이 없는 것(예로, X)으로 식별되며, TV와는 관련도가 중간인 것(예로, △)으로 식별될 수 있으며, 스피커와는 관련이 있는 것(예로, 0)으로 식별될 수 있다. 그리고, 전자 장치(100)는 각각의 키워드와 관련도가 가장 높은 스피커를 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치로 식별할 수 있다.
- [0089] 그리고, 도 7b를 참조하면, 최종 텍스트 정보가 “아이유 뮤직비디오 들어줘” 인 경우, 도 7a의 최종 텍스트 정보와 비교할 때, “뮤직비디오” 에 대한 키워드가 상이하며, 전자 장치(100)는 “뮤직비디오” 와 관련도가 높은 TV를 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치로 식별할 수 있다.
- [0090] 그리고, 도 7c를 참조하면, 최종 텍스트 정보가 “백종원 목은지 김치찌개 레시피 알려줘” 인 경우, 전자 장치(100)는 최종 텍스트 정보로부터 “백종원”, “목은지”, “김치찌개”, “레시피” 및 “알려줘” 의 5개의 키워드를 식별할 수 있다. 그리고, 전자 장치(100)는 5개의 키워드를 바탕으로, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치를 냉장고로 식별할 수 있다.
- [0091] 도 7a 내지 도 7c에서는 최종 텍스트 정보에 대응되는 사용자가 발화한 음성의 음절 각각에 대응되는 키워드 모두를 이용하여 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치가 식별되는 것으로 도시하였으나, 이에 한정되지 않고, 최종 텍스트 정보에 대응되는 사용자가 발화한 음성의 음절 각각에 대응되는 키워드 중 적어도 하나를 통해서도 사용자가 발화한 음성에 대응되는 동작을 수행할 외부 장치가 식별될 수 있다.
- [0092] 상술한 실시 예를 통해, 전자 장치(100)는 복수의 외부 장치 중 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치를 결정함으로써, 댁내의 복수의 외부 장치가 사용자 발화 음성에 의해 동시에 응답하게 되는 문제점을

방지할 수 있다.

- [0094] 도 8은 본 개시의 일 실시 예에 따른 전자 장치 및 복수의 외부 장치간의 동작을 설명하기 위한 시퀀스도이다.
- [0095] 우선, 도 8을 참조하면, 전자 장치(100)는 트리거 음성을 수신할 수 있다(S805). 일 실시 예로, 트리거 음성은 음성인 “빅스비”, “하이 빅스비” 등과 같은 트리거 단어를 포함할 수 있다.
- [0096] 그리고, 트리거 음성이 수신되면, 전자 장치(100)는 음성 인식 모드로 진입할 수 있다(S810). 구체적으로, 전자 장치(100)는 음성 인식 모드로 진입하여 인공지능 에이전트 프로그램을 활성화 할 수 있다. 또한, 트리거 음성이 수신되면, 전자 장치(100)는 제1 외부 장치(10) 및 제2 외부 장치(20)로 음성 인식 모드로 진입 위한 정보를 전송할 수 있다(S815).
- [0097] 그리고, 전자 장치(100)는 음성 인식 모드로 동작하는 도중 사용자 발화 음성에 대응되는 제1 음성 정보를 획득할 수 있다(S830). 또한, 전자 장치(100)로부터 음성 인식 모드로 진입 위한 정보를 수신한 제1 외부 장치(10)는 음성 인식 모드로 진입하여(S820) 사용자가 발화한 음성에 대응되는 제2-1 음성 정보를 획득할 수 있다(S835). 그리고, 전자 장치(100)로부터 음성 인식 모드로 진입 위한 정보를 수신한 제2 외부 장치(20)는 음성 인식 모드로 진입하여(S825) 사용자가 발화한 음성에 대응되는 제2-2 음성 정보를 획득할 수 있다(S840).
- [0098] 그리고, 제1 외부 장치(10)는 전자 장치(100)로 제2-1 음성 정보를 전송하고(S845), 제2 외부 장치(20)는 전자 장치(100)로 제2-2 음성 정보를 전송할 수 있다(S850). 전자 장치(100)는 제2-1 음성 정보 및 제2-2 음성 정보를 수신한 경우, 사용자 발화 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다(S855). 구체적으로, 전자 장치(100)는 제1 음성 정보, 제2-1 음성 정보 및 제2-2 음성 정보를 바탕으로 사용자 발화 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다.
- [0099] 그리고, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 제1 외부 장치(10)인 것으로 식별되면, 전자 장치(100)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 제1 외부 장치(10)로 전송할 수 있다(S860). 그리고, 제1 외부 장치(10)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 바탕으로, 사용자 발화 음성에 대응되는 동작을 수행할 수 있다(S865).
- [0100] 다만, 이에 한정되지 않으며, 그리고, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 제2 외부 장치(20)인 것으로 식별되면, 전자 장치(100)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 제2 외부 장치(20)로 전송할 수 있다. 또한, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 전자 장치(100)인 것으로 식별되면, 전자 장치(100)는 사용자 발화 음성에 대응되는 동작을 수행할 수 있다.
- [0102] 도 9은 본 개시의 일 실시 예에 따른 전자 장치 및 복수의 외부 장치간의 동작을 설명하기 위한 시퀀스도이다.
- [0103] 우선, 도 9을 참조하면, 제1 외부 장치(10)는 트리거 음성을 수신할 수 있다(S905). 그리고, 트리거 음성이 수신되면, 제1 외부 장치(10)는 음성 인식 모드로 진입할 수 있다(S910). 구체적으로, 제1 외부 장치(10)는 음성 인식 모드로 진입하여 인공지능 에이전트 프로그램을 활성화 할 수 있다. 또한, 트리거 음성이 수신되면, 제1 외부 장치(10)는 전자 장치(100) 및 제2 외부 장치(20)로 음성 인식 모드로 진입 위한 정보를 전송할 수 있다(S915).
- [0104] 그리고, 제1 외부 장치(10)는 음성 인식 모드로 동작하는 도중 사용자 발화 음성에 대응되는 제2-1 음성 정보를 획득할 수 있다(S930). 또한, 제1 외부 장치(10)로부터 음성 인식 모드로 진입 위한 정보를 수신한 전자 장치(100)는 음성 인식 모드로 진입하여(S920) 사용자가 발화한 음성에 대응되는 제1 음성 정보를 획득할 수 있다(S935). 그리고, 제1 외부 장치(10)로부터 음성 인식 모드로 진입 위한 정보를 수신한 제2 외부 장치(20)는 음성 인식 모드로 진입하여(S925) 사용자가 발화한 음성에 대응되는 제2-2 음성 정보를 획득할 수 있다(S940).
- [0105] 그리고, 제1 외부 장치(10)는 전자 장치(100)로 제2-1 음성 정보를 전송하고(S945), 제2 외부 장치(20)는 전자 장치(100)로 제2-2 음성 정보를 전송할 수 있다(S950). 전자 장치(100)는 제2-1 음성 정보 및 제2-2 음성 정보를 수신한 경우, 사용자 발화 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다(S955). 구체적으로, 전자 장치(100)는 제1 음성 정보, 제2-1 음성 정보 및 제2-2 음성 정보를 바탕으로 사용자 발화 음성에 대응되는 동작 및 동작을 수행할 외부 장치를 식별할 수 있다.
- [0106] 그리고, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 제1 외부 장치(10)인 것으로 식별되면, 전자

장치(100)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 제1 외부 장치(10)로 전송할 수 있다(S960). 그리고, 제1 외부 장치(10)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 바탕으로, 사용자 발화 음성에 대응되는 동작을 수행할 수 있다(S965).

- [0107] 다만, 이에 한정되지 않으며, 그리고, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 제2 외부 장치(20)인 것으로 식별되면, 전자 장치(100)는 사용자 발화 음성에 대응되는 동작에 대한 정보를 제2 외부 장치(20)로 전송할 수 있다. 또한, 사용자 발화 음성에 대응되는 동작을 수행할 외부 장치가 전자 장치(100)인 것으로 식별되면, 전자 장치(100)는 사용자 발화 음성에 대응되는 동작을 수행할 수 있다.
- [0108] 상술한 바와 같이 본 개시의 다양한 실시 예를 도면을 통해 설명하였다. 그러나, 이는 본 개시의 기술을 특정할 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 실시 예들의 다양한 변경(modifications), 균등물(equivalents), 및/또는 대체물(alternatives)을 포함하는 것으로 이해되어야 한다.
- [0109] 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.
- [0110] 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.
- [0111] 본 개시에서 사용된 "제 1," "제 2," "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. 예를 들면, 제 1 사용자 기기와 제 2 사용자 기기는, 순서 또는 중요도와 무관하게, 서로 다른 사용자 기기를 나타낼 수 있다. 예를 들면, 본 개시에 기재된 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 바꾸어 명명될 수 있다.
- [0112] 본 개시에서 사용된 "모듈", "유닛", "부(part)" 등과 같은 용어는 적어도 하나의 기능이나 동작을 수행하는 구성요소를 지칭하기 위한 용어이며, 이러한 구성요소는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다. 또한, 복수의 "모듈", "유닛", "부(part)" 등은 각각이 개별적인 특정한 하드웨어로 구현될 필요가 있는 경우를 제외하고는, 적어도 하나의 모듈이나 칩으로 일체화되어 적어도 하나의 프로세서로 구현될 수 있다. 그리고, 그리고, 사용자라는 용어는 전자 장치를 사용하는 사람 또는 전자 장치를 사용하는 장치(예: 인공지능 전자 장치)를 지칭할 수 있다.
- [0113] 어떤 구성요소(예: 제 1 구성요소)가 다른 구성요소(예: 제 2 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: 제 3 구성요소)를 통하여 연결될 수 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소(예: 제 1 구성요소)가 다른 구성요소(예: 제 2 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.
- [0114] 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) 프로세서"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.
- [0115] 본 개시에서 사용된 용어들은 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 다른 실시예의 범위를 한정하려는 의도가 아닐 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 용어들은 본 개시에 기재된 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가질 수 있다. 본 개시에 사용된 용어들 중

일반적인 사전에 정의된 용어들은, 관련 기술의 문맥상 가지는 의미와 동일 또는 유사한 의미로 해석될 수 있으며, 본 개시에서 명백하게 정의되지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다. 경우에 따라서, 본 개시에서 정의된 용어일지라도 본 개시의 실시예들을 배제하도록 해석될 수 없다.

[0116] 본 개시의 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(100))를 포함할 수 있다. 명령이 프로세서에 의해 실행될 경우, 프로세서가 직접, 또는 상기 프로세서의 제어 하에 다른 구성요소들을 이용하여 상기 명령에 해당하는 기능을 수행할 수 있다. 명령은 컴파일러 또는 인터프리터에 의해 생성 또는 실행되는 코드를 포함할 수 있다.

[0117] 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적 저장매체'는 실재(tangible)하는 장치이고, 신호(signal)(예를 들어, 전자기파)를 포함하지 않는다는 것을 의미할 뿐, 이용어가 데이터가 저장매체에 반영구적 또는 임시적으로 저장됨을 구분하지 않는다. 예로, '비일시적 저장매체'는 데이터가 임시적으로 저장되는 버퍼를 포함할 수 있다.

[0118] 다양한 실시 예들에 따른 구성 요소(예: 모듈 또는 프로그램) 각각은 단수 또는 복수의 개체로 구성될 수 있으며, 전술한 해당 서브 구성 요소들 중 일부 서브 구성 요소가 생략되거나, 또는 다른 서브 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.

[0119] 일 실시예에 따르면, 본 개시에 개시된 다양한 실시예들에 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 배포되거나, 또는 어플리케이션 스토어(예: 플레이 스토어™)를 통해 또는 두개의 사용자 장치들(예: 스마트폰들) 간에 직접, 온라인으로 배포(예: 다운로드 또는 업로드)될 수 있다. 온라인 배포의 경우에, 컴퓨터 프로그램 제품(예: 다운로드를 앱(downloadable app))의 적어도 일부는 제조사의 서버, 어플리케이션 스토어의 서버, 또는 중계 서버의 메모리와 같은 기기로 읽을 수 있는 저장 매체에 적어도 일시 저장되거나, 임시적으로 생성될 수 있다.

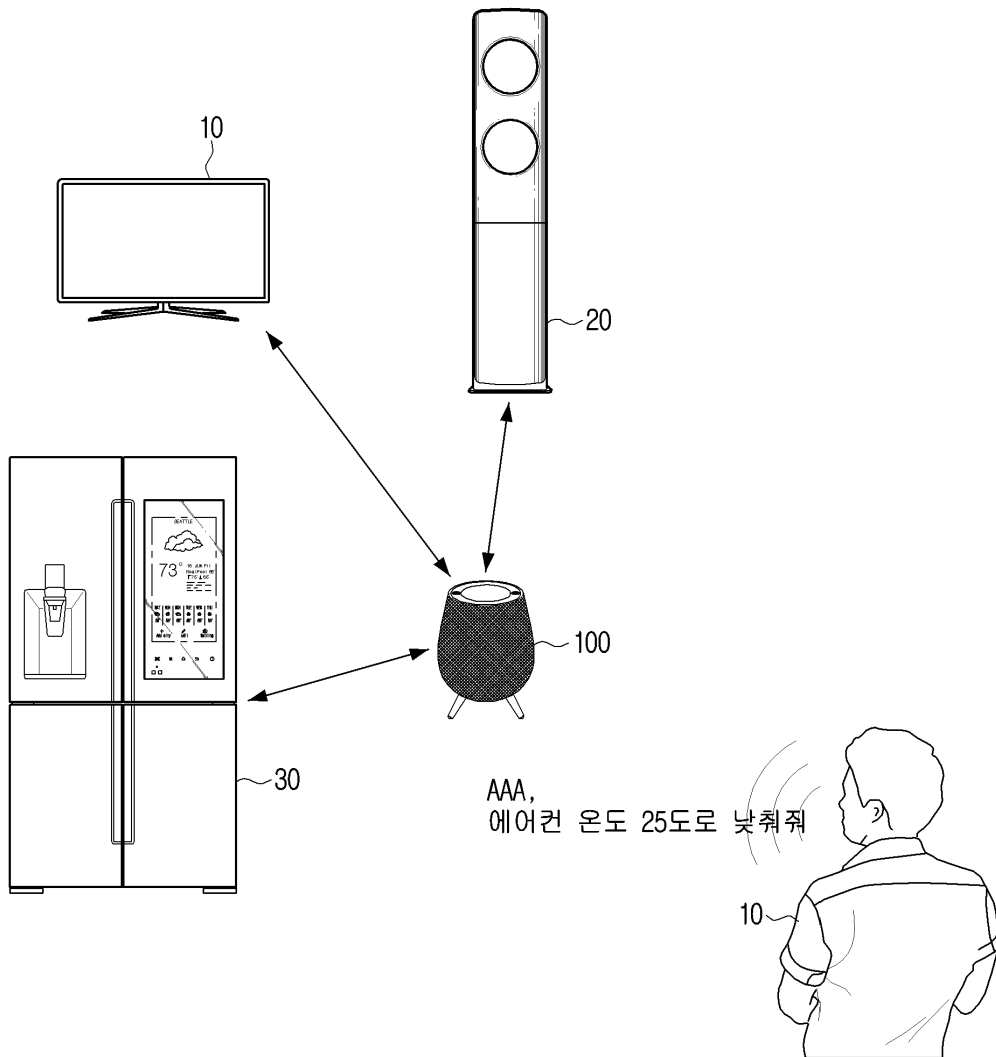
[0120] 또한, 이상에서는 본 개시의 바람직한 실시 예에 대하여 도시하고 설명하였지만, 본 개시는 상술한 특정의 실시 예에 한정되지 아니하며, 청구범위에서 청구하는 본 개시의 요지를 벗어남이 없이 당해 개시가 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안 될 것이다.

부호의 설명

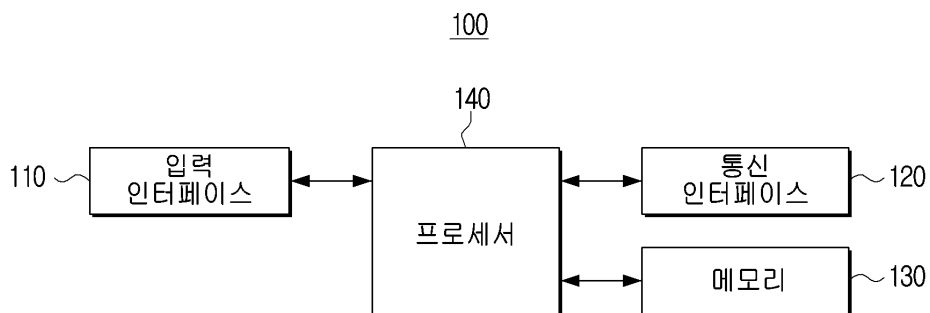
- [0122] 100: 전자 장치
- 110: 입력 인터페이스
- 120: 통신 인터페이스
- 130: 메모리

도면

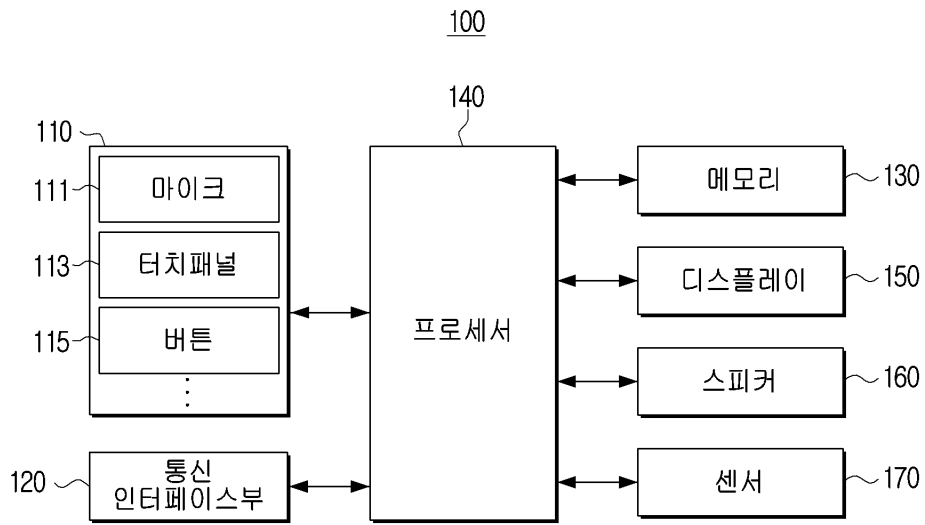
도면1



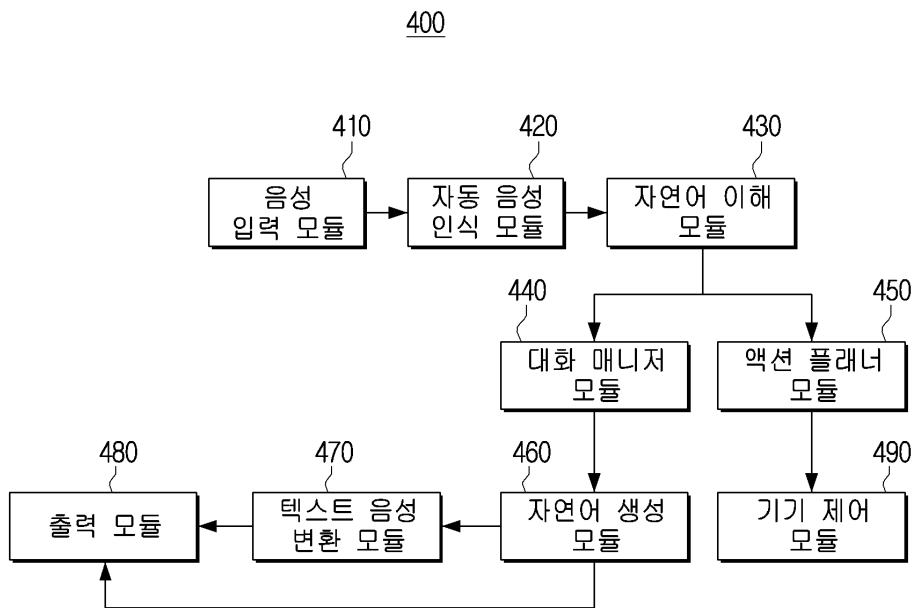
도면2



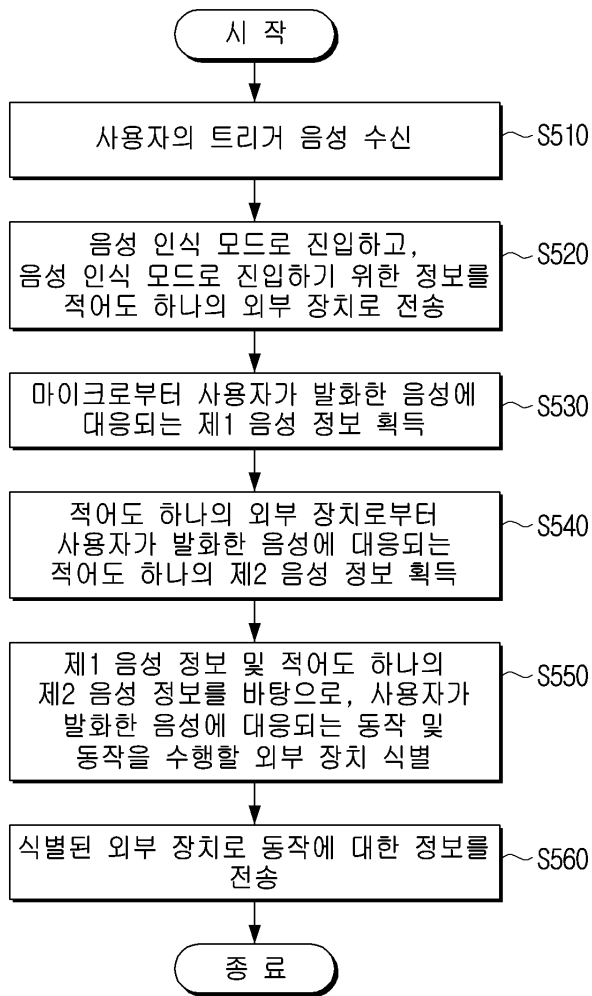
도면3



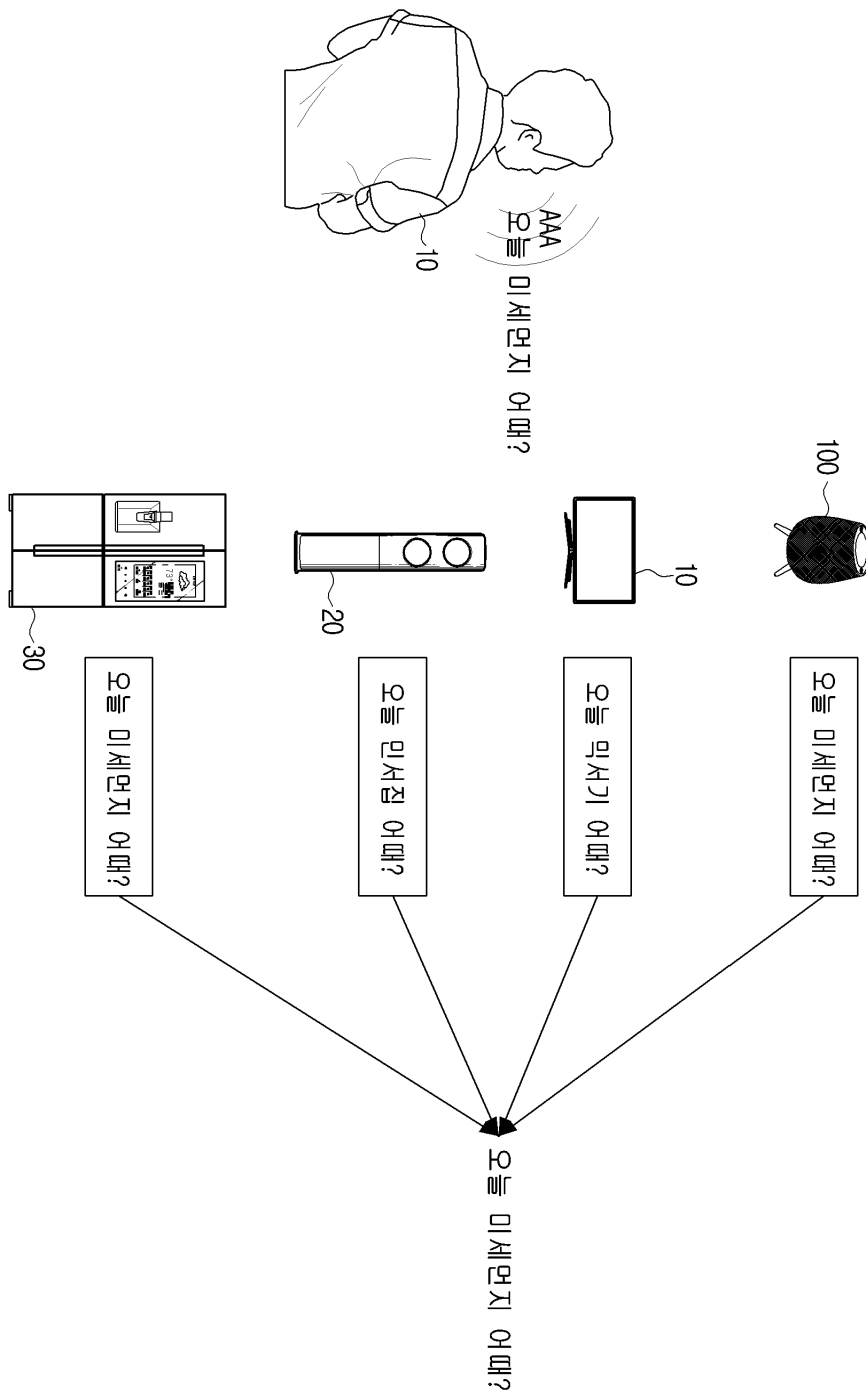
도면4



도면5



도면6



도면7a

	"아이유"	"노래"	"틀어줘"
냉장고	X	X	△
TV	○	△	○
Speaker	○	○	○

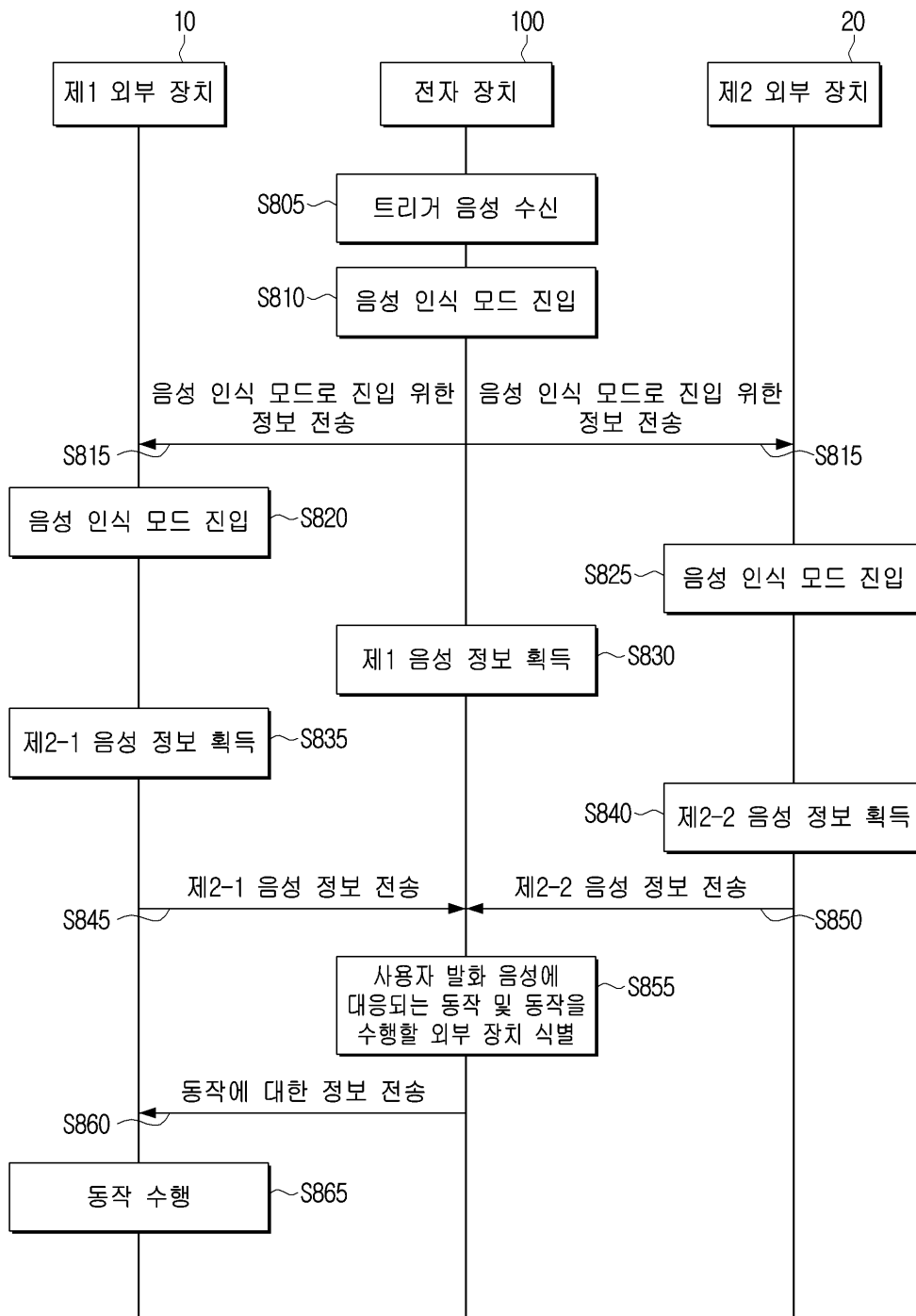
도면7b

	"아이유"	"뮤직비디오"	"틀어줘"
냉장고	X	X	△
TV	○	○	○
Speaker	○	△	○

도면7c

	"백종원"	"묵은지"	"김치찌개"	"레시피"	"알려줘"
냉장고	○	○	○	○	△
TV	○	X	X	X	△
Speaker	△	X	X	X	△

도면8



도면9

