



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0045851
(43) 공개일자 2020년05월06일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) *G06F 3/041* (2006.01)
G06F 3/048 (2017.01) *G06F 3/16* (2018.01)
G10L 15/02 (2006.01) *G10L 15/04* (2006.01)
G10L 15/183 (2013.01) *G10L 15/26* (2006.01)
 (52) CPC특허분류
G10L 15/22 (2013.01)
G06F 3/041 (2013.01)
 (21) 출원번호 10-2018-0126894
 (22) 출원일자 2018년10월23일
 심사청구일자 없음

(71) 출원인
삼성전자주식회사
 경기도 수원시 영통구 삼성로 129 (매탄동)
 (72) 발명자
한지현
 서울특별시 강남구 강남대로146길 28, 논현아파트
 103-407
안유현
 경상남도 창원시 성산구 원이대로 774, 성원아파트
 307-801
 (74) 대리인
특허법인태평양

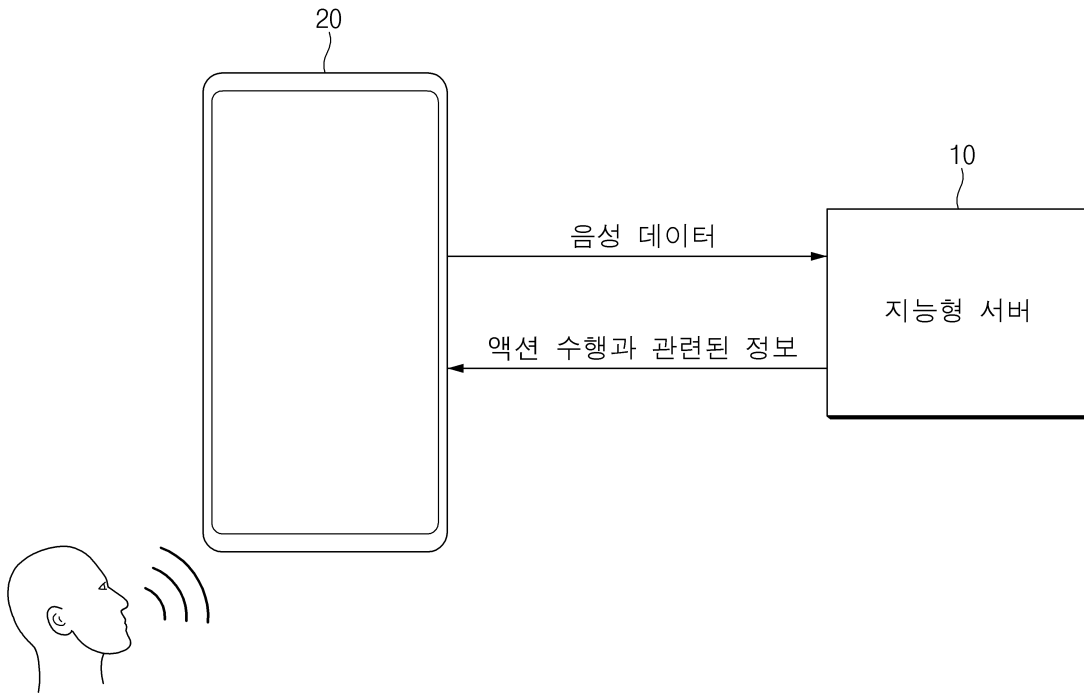
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **음성 인식 서비스를 제공하는 전자 장치 및 시스템**

(57) 요약

일 실시예에 따른 시스템에 포함된 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 프로세서로 하여금, 마이크를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 웨이크업(wake-up) 발화(utterance)를 수신하고, 상기 웨이크업 발화 이후에 마이크를 통해, 제1 사용자 발화(뒷면에 계속)

대표도 - 도1



를 수신하고, 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고, 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고, 상기 웨이크업 발화 수신 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고, 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고, 상기 제2 사용자 발화에 대한 텍스트에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상기 제2 사용자 발화에 대한 텍스트에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장할 수 있다.

(52) CPC특허분류

G06F 3/048 (2013.01)

G06F 3/16 (2019.02)

G10L 15/02 (2013.01)

G10L 15/04 (2013.01)

G10L 15/183 (2013.01)

G10L 15/26 (2013.01)

G10L 2015/223 (2013.01)

명세서

청구범위

청구항 1

시스템에 있어서,

사용자 인터페이스;

마이크;

스피커;

상기 사용자 인터페이스, 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서; 및

상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고,

상기 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금,

상기 마이크를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 웨이크업(wake-up) 발화(utterance)를 수신하고,

상기 웨이크업 발화 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고,

상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고,

상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고,

상기 웨이크업 발화 수신 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고,

상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고,

상기 제2 사용자 발화에 대한 텍스트에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고,

상기 제2 사용자 발화에 대한 텍스트에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장하는 시스템.

청구항 2

청구항 1에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 선택된 단어 또는 구문을 수신하도록 구성된 상기 사용자 인터페이스를 제공하거나, 상기 프로세서의 동작에 의해 상기 선택된 단어 또는 구문을 결정하도록 하는 시스템.

청구항 3

청구항 1에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 선택된 시간 구간을 수신하도록 구성된 상기 사용자 인터페이스를 제공하도록 하는 시스템.

청구항 4

청구항 1에 있어서, 상기 선택된 시간은,

상기 제1 응답이 생성된 이후부터의 시간을 포함하는 시스템.

청구항 5

청구항 1에 있어서, 상기 선택된 단어 또는 구문은,

카테고리에 관련된 단어를 포함하는 시스템.

청구항 6

청구항 1에 있어서, 상기 인스트럭션들은, 상기 프로세서가,
 상기 제1 사용자 발화에 기반하여 발화 속도 변화, 톤 변화 및 억양 변화를 확인하고,
 상기 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 지정된 조건을 만족하는 경우, 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 포함되는지를 확인하도록 하는 시스템.

청구항 7

청구항 1에 있어서, 상기 인스트럭션들은, 상기 프로세서가,
 상기 제2 사용자 발화에, 상기 선택된 단어 또는 구문과 상이한 다른 선택된 단어 또는 구문이 포함되는지를 확인하고,
 상기 제2 사용자 발화에, 상기 다른 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고,
 상기 제2 사용자 발화에, 상기 다른 선택된 단어 또는 구문이 포함되어 있으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 시스템.

청구항 8

청구항 1에 있어서,
 상기 인스트럭션들은, 상기 프로세서가,
 상기 선택된 단어 또는 구문을 포함하는 문장에서 상기 선택된 단어 또는 구문의 위치에 기반하여, 상기 문장의 오인식 여부를 확인하고,
 상기 문장이 오인식된 경우, 상기 선택된 단어 또는 구문을 포함하는 문장에 대한 재발화를 요청하는 소리를 상기 스피커를 통해 출력하도록 하는 시스템.

청구항 9

시스템에 있어서,
 사용자 인터페이스;
 마이크;
 스피커;
 상기 사용자 인터페이스, 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서; 및
 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고,
 상기 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금,
 상기 사용자 인터페이스를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 사용자 입력을 수신하고,
 상기 사용자 입력을 수신한 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고,
 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고,
 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고,
 상기 사용자 입력을 수신한 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고,
 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고,
 상기 제2 사용자 발화에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고,

상기 제2 사용자 발화에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장하는 시스템.

청구항 10

청구항 9에 있어서, 상기 사용자 인터페이스는,

터치스크린 디스플레이, 프레스(press) 가능한 버튼, 또는 터치 가능한 버튼 중 적어도 하나를 포함하는 시스템.

청구항 11

청구항 9에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 선택된 단어 또는 구문을 수신하도록 구성된 상기 사용자 인터페이스를 제공하거나, 상기 프로세서의 동작에 의해 상기 선택된 단어 또는 구문을 결정하도록 하는 시스템.

청구항 12

전자 장치에 있어서,

통신 회로;

마이크; 및

상기 통신 회로, 상기 입력 회로 및 상기 마이크와 기능적으로 연결된 프로세서; 및

상기 프로세서와 기능적으로 연결된 메모리를 포함하고,

상기 메모리는, 실행될 때에 상기 프로세서로 하여금,

상기 마이크를 통해 음성 인식 서비스를 호출하는 웨이크업 발화를 수신하면, 상기 음성 인식 서비스를 제공할 수 있는 지능형 앱을 실행하고,

상기 지능형 앱을 이용하여, 상기 마이크를 통해 제1 사용자 발화를 수신하고,

상기 지능형 앱을 이용하여, 상기 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행하고,

상기 제1 액션을 수행한 시점으로부터 선택된 시간 내에 상기 마이크를 통해 제2 사용자 발화를 수신하고

상기 지능형 앱을 이용하여, 선택된 시간상기 제2 사용자 발화에 기반하여 선택된 단어 또는 구문이 인식되는지를 확인하고,

상기 선택된 시간 내에 상기 제2 사용자 발화에 기반하여 상기 선택된 단어 또는 구문이 인식되면, 상기 지능형 앱을 이용하여, 상기 제2 사용자 발화를 기반으로 결정된 제2 액션을 수행하고,

상기 선택된 시간 내에 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 인식되지 않으면, 상기 지능형 앱을 종료하도록 설정된 인스트럭션들을 저장하는 전자 장치.

청구항 13

청구항 12에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 제1 액션을 수행한 시점으로부터 상기 선택된 시간 내에 상기 선택된 단어 또는 구문이 인식되면, 상기 선택된 단어 또는 구문을 포함하는 문장을 결정하고,

상기 문장을 상기 통신 회로를 통해 외부 전자 장치로 송신하고,

상기 외부 전자 장치로부터 상기 문장에 기반하여 결정된 제2 액션 수행과 관련된 정보를 수신하고,

상기 제2 액션 수행과 관련된 정보에 기반하여 상기 제2 액션을 수행하도록 더 설정된 전자 장치.

청구항 14

청구항 13에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 문장에 다른 선택된 단어 또는 구문이 포함되는지를 확인하고,

상기 문장에 상기 다른 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 문장을 상기 통신 회로를 통해 상기 외부 전자 장치로 송신하고,

상기 문장에 상기 다른 선택된 단어 또는 구문이 포함되어 있으면, 상기 문장을 상기 외부 전자 장치로 송신하지 않도록 더 설정된 전자 장치.

청구항 15

청구항 13에 있어서,

스피커를 더 포함하고, 상기 인스트럭션들은, 상기 프로세서가,

상기 문장에서 상기 선택된 단어 또는 구문의 위치에 기반하여, 상기 확인된 문장의 오인식 여부를 확인하고,

상기 문장의 오인식된 경우, 상기 선택된 단어 또는 구문을 포함하는 문장에 대한 재발화를 요청하는 소리를 상기 스피커를 통해 출력하도록 더 설정된 전자 장치.

청구항 16

청구항 12에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 제2 사용자 발화에 기반하여 또 다른 선택된 단어 또는 구문이 인식되면, 상기 지능형 앱을 종료하도록 더 설정된 전자 장치.

청구항 17

청구항 12에 있어서, 상기 선택된 단어 또는 구문은,

상기 전자 장치에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로를 통해 지정된 액션 요청과 관련된 단어를 포함하는 전자 장치.

청구항 18

청구항 12에 있어서, 상기 선택된 단어 또는 구문은,

상기 전자 장치에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로를 통해 지정된 카테고리에 속하는 액션 요청과 관련된 단어를 포함하는 전자 장치.

청구항 19

청구항 12에 있어서, 상기 선택된 단어 또는 구문은,

상기 전자 장치에서 수행 가능한 복수의 액션들을 요청하는 단어와, 화제 전환을 위한 단어 및 상기 전자 장치를 지칭하는 단어 중 적어도 하나를 더 포함하는 전자 장치.

청구항 20

청구항 12에 있어서, 상기 인스트럭션들은, 상기 프로세서가,

상기 제1 사용자 발화에 기반하여 발화 속도 변화, 톤 변화 및 억양 변화를 확인하고,

상기 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 지정된 조건을 만족하는 경우, 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 포함되는지를 확인하도록 더 설정된 전자 장치.

발명의 설명

기술 분야

[0001] 본 문서에서 개시되는 실시예들은, 음성 인식 서비스 제공 기술과 관련된다.

배경 기술

[0002] 전자 장치가 다기능화 및 고성능화됨에 따라 전자 장치에 음성 인식 기술이 적용되는 사례가 늘고 있다. 음성 인식 기술이 적용된 전자 장치는 사용자 발화에 기반하여 사용자의 의도(기능 요청)를 확인하고, 사용자의 의도에 따른 응답을 제공할 수 있다. 전자 장치는 웨이크업 발화(예: “하이 빅스비”) 또는 버튼 입력을 수신하면, 음성 인식 서비스를 제공하는 지정된 어플리케이션(지능형 앱) 또는 인텔리전트 어시스턴스 서비스를 실행하고, 지정된 어플리케이션을 통해 사용자 발화를 수신하고, 사용자 발화에 기반하여 사용자의 의도에 따른 하나의 응답을 제공한 후 다시 지정된 어플리케이션을 종료하였다. 이후, 전자 장치는 다시 웨이크업 발화 또는 버튼 입력을 수신하여야 음성 인식 서비스를 제공할 수 있었다.

발명의 내용

해결하려는 과제

[0003] 따라서, 사용자는 음성 인식 서비스를 이용하여 복수의 기능을 요청하기 위해서는 웨이크업 발화(또는, 버튼 입력)와 사용자 발화를 번갈아 반복적으로 수행해야 하는 불편이 있었다.

[0004] 본 문서에 개시되는 다양한 실시예들은 한 번의 웨이크업 발화 후에 복수의 사용자 발화들에 각기 대응하는 복수의 응답들을 제공할 수 있는 음성 인식 서비스를 제공하는 전자 장치 및 시스템을 제공한다.

과제의 해결 수단

[0005] 본 문서에 개시되는 일 실시예에 따른 시스템은, 마이크; 스피커; 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서; 및 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고, 상기 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 상기 마이크를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 웨이크업(wake-up) 발화(utterance)를 수신하고, 상기 웨이크업 발화 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고, 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고, 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고, 상기 웨이크업 발화 수신 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고, 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고, 상기 제2 사용자 발화에 대한 텍스트에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상기 제2 사용자 발화에 대한 텍스트에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장할 수 있다.

[0006] 또한, 본 문서에 개시되는 일 실시예에 따른 시스템은, 사용자 인터페이스; 마이크; 스피커; 상기 사용자 인터페이스, 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서; 및 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고, 상기 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 상기 사용자 인터페이스를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 사용자 입력을 수신하고, 상기 사용자 입력을 수신한 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고, 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고, 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고, 상기 사용자 입력을 수신한 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고, 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고, 상기 제2 사용자 발화에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상기 제2 사용자 발화에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장할 수 있다.

[0007] 또한, 본 문서에 개시되는 일 실시예에 따른 전자 장치는, 통신 회로; 마이크; 및 상기 통신 회로, 상기 입력 회로 및 상기 마이크와 기능적으로 연결된 프로세서; 및 상기 프로세서와 기능적으로 연결된 메모리를 포함하고, 상기 메모리는, 실행될 때에 상기 프로세서로 하여금, 상기 마이크를 통해 음성 인식 서비스를 호출하는 웨이크업 발화를 수신하면, 상기 음성 인식 서비스를 제공할 수 있는 지능형 앱을 실행하고, 상기 지능형 앱을 이용하여, 상기 마이크를 통해 제1 사용자 발화를 수신하고, 상기 지능형 앱을 이용하여, 상기 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행하고, 상기 제1 액션을 수행한 시점으로부터 선택된 시간 내에 상기 마이크를 통해 제2 사용자 발화를 수신하고 상기 지능형 앱을 이용하여, 상기 제2 사용자 발화에 기반하여 선택된 단어 또는 구문이 인식되는지를 확인하고, 상기 선택된 시간 내에 상기 제2 사용자 발화에 기반하여 상기 선택된 단어 또는 구문이 인식되면, 상기 지능형 앱을 이용하여, 상기 제2 사용자 발화를 기반으로 결정된 제2 액션

을 수행하고, 상기 선택된 시간 내에 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 인식되지 않으면, 상기 지능형 앱을 종료하도록 설정된 인스트럭션들을 저장할 수 있다.

발명의 효과

[0008] 본 문서에 개시되는 실시예들에 따르면, 한 번의 웨이크업 발화 후에 복수의 사용자 발화들에 각기 대응하는 복수의 응답들을 제공할 수 있다. 이 외에, 본 문서를 통해 직접적 또는 간접적으로 파악되는 다양한 효과들이 제공될 수 있다.

도면의 간단한 설명

[0009] 도 1은 일 실시예에 따른 통합 지능 시스템의 구성도를 나타낸다.
 도 2는 일 실시예에 따른 전자 장치의 구성도를 나타낸다.
 도 3a는 일 실시예에 따른 대화 모드가 사용 설정되지 않은 경우의 전자 장치의 음성 인식 서비스를 위한 동작 모드들을 나타낸다.
 도 3b는 일 실시예에 따른 대화 모드가 사용 설정되지 않은 경우의 음성 인식 서비스의 예를 나타낸다.
 도 4a는 일 실시예에 따른 대화 모드가 사용 설정된 경우의 전자 장치의 음성 인식 서비스를 위한 동작 모드들을 나타낸다.
 도 4b는 일 실시예에 따른 대화 모드가 사용 설정된 경우의 음성 인식 서비스의 예를 나타낸다.
 도 5a는 일 실시예에 따른 대화 모드의 사용을 설정하기 위한 UI 화면을 나타낸다.
 도 5b는 일 실시예에 따른 대화 모드에 대한 시간 설정을 위한 UI 화면을 나타낸다.
 도 5c는 일 실시예에 따른 대화 모드에 대한 트리거 단어 설정을 위한 UI 화면의 일 예를 나타낸다.
 도 5d는 일 실시예에 따른 대화 모드에 대한 트리거 단어 설정을 위한 UI 화면의 다른 예를 나타낸다.
 도 6은 일 실시예에 따른 프로세서의 구성요소들의 일 예를 나타낸다.
 도 7는 일 실시예에 따른 프로세서의 구성요소들의 다른 예를 나타낸다.
 도 8은 일 실시예에 따른 통합 지능 시스템에 의한 음성 인식 서비스 제공 방법을 나타낸다.
 도 9는 일 실시 예에 따른 전자 장치에 의한 음성 인식 서비스 제공 방법을 나타낸다.
 도 10은 다양한 실시예에 따른 네트워크 환경 내의 전자 장치의 블록도를 나타낸다.
 도 11은 일 실시예에 따른 통합 지능 (integrated intelligence) 시스템을 나타낸 블록도이다.
 도 12는 일 실시예에 따른, 컨셉과 액션의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.
 도 13은 일 실시예에 따라, 지능형 앱을 통해 수신된 음성 입력을 처리하는 화면을 표시하는 사용자 단말을 도시하는 도면이다.
 도면의 설명과 관련하여, 동일 또는 유사한 구성요소에 대해서는 동일 또는 유사한 참조 부호가 사용될 수 있다.

발명을 실시하기 위한 구체적인 내용

[0010] 도 1은 일 실시예에 따른 통합 지능 시스템의 구성도를 나타낸다.
 [0011] 도 1을 참조하면, 일 실시예에 따르면, 전자 장치(20)는 마이크를 통해 웨이크업 발화(예: 하이 빅스비)를 수신하면, 음성 인식 서비스를 제공하는 지능형 앱(또는, 지능형 에이전트) 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다. 상기 웨이크업 발화는 특정 단어, 구문 또는 문장을 포함하고, 특정 사용자의 음향 특성(예: 특정 주파수 특성)을 갖는 발화일 수 있다. 다양한 실시 예에 따르면, 전자 장치(20)가 입력 회로(예: 터치 센서 또는 물리적 버튼)를 포함하면, 전자 장치(20)는 입력 회로를 통해 지정된 입력(예: 터치 또는 눌림)을 확인하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다.
 [0012] 일 실시예에 따르면, 전자 장치(20)의 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 마이크를 통해 제1 사용

자 발화(voice input)를 수신하면, 제1 사용자 발화에 기반하여 사용자의 의도에 따른 액션을 수행할 수 있다. 예를 들어, 전자 장치(20)의 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 제1 사용자 발화를 수신하면, 제1 사용자 발화를 제1 음성 데이터(예: PCM(pulse code modulation) 데이터)로 변환하고, 제1 음성 데이터를 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 제1 음성 데이터를 수신하면, 제1 음성 데이터를 텍스트 데이터로 변환하고, 변환된 텍스트 데이터를 기반으로 사용자의 의도를 결정할 수 있다. 지능형 서버(10)는 결정된 사용자의 의도에 따른 제1 액션을 결정하고, 결정된 제1 액션 수행과 관련된 정보를 생성하고 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 상기 제1 액션 수행과 관련된 정보는 예를 들면, 제1 액션을 수행하는 애플리케이션의 정보 및 해당 애플리케이션이 수행하는 기능 정보를 포함할 수 있다. 전자 장치(20)는 지능형 서버(10)로부터 제1 액션 수행과 관련된 정보를 수신하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스에 의해 제1 액션 수행과 관련된 정보에 기반하여 제1 사용자 발화에 대응하는 제1 액션을 수행할 수 있다. 전자 장치(20)는 제1 액션의 수행 과정에서 또는 제1 액션의 수행 완료 시에 제1 액션 수행과 관련된 제1 소리를 출력할 수 있다. 상기 제1 액션 수행과 관련된 제1 소리는 예를 들면, 지능형 서버(10)로부터 제공된 소리가거나, 전자 장치(20)가 제1 액션 수행과 관련된 정보를 기반으로 생성한 소리일 수 있다. 상기 제1 액션 수행과 관련된 제1 소리는 예를 들면, 제1 액션 수행 과정을 안내하는 소리 또는 제1 액션 수행의 결과를 안내하는 소리일 수 있다.

[0013] 일 실시예에 따르면, 전자 장치(20)의 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 마이크를 통해 수신된 제2 사용자 발화에 기반하여 트리거 단어가 인식되는지를 확인할 수 있다. 전자 장치(20)의 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 수신된 제2 사용자 발화에 기반하여 트리거 단어가 인식되면, 트리거 단어를 포함하는 문장에 기반하여 결정된 제2 액션을 수행할 수 있다. 반면, 전자 장치(20)의 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제2 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 종료될 수 있다.

[0014] 상술한 실시예에 따르면, 전자 장치(20)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행시키기 위한(또는, 음성 인식 서비스를 호출하기 위한) 한 번의 시도(예: 웨이크업 발화 또는 지정된 입력)를 통해 복수의 사용자 발화들에 각기 대응하는 각기 복수의 액션들을 수행할 수 있다.

[0016] 도 2는 일 실시예에 따른 전자 장치의 구성도를 나타낸다.

[0017] 도 2을 참조하면, 전자 장치(20)는 입력 회로(210), 마이크(220), 통신 회로(230), 스피커(240), 메모리(250) 및 프로세서(260)를 포함할 수 있다. 일 실시예에서, 전자 장치(20)는 일부 구성요소가 생략되거나, 추가적인 구성요소를 더 포함할 수 있다. 예를 들어, 전자 장치(20)는 입력 회로(210)를 포함하지 않을 수 있다. 일 실시예에서, 전자 장치(20)의 구성요소들 중 일부가 결합되어 하나의 개체로 구성되며, 결합 이전의 해당 구성요소들의 기능을 동일하게 수행할 수 있다.

[0018] 일 실시예에 따르면, 입력 회로(210)는 터치스크린 디스플레이, 프레스 가능한 버튼 또는 터치 가능한 버튼 중 적어도 하나를 포함할 수 있다. 상기 프레스 가능한 버튼 또는 터치 가능한 버튼은 예를 들면, 전자 장치(20)의 전면 또는 측면에 구비될 수 있다. 입력 회로(210)가 터치스크린 디스플레이를 포함하면, 전자 장치(20)는 터치스크린 디스플레이를 통해 디스플레이 기능을 제공할 수 있다.

[0019] 일 실시예에 따르면, 마이크(220)는 사용자 발화를 수신할 수 있다. 예를 들어, 마이크(220)는 사용자 발화를 수신하고, 수신된 사용자 발화에 대응하는 아날로그 또는 디지털 신호를 출력할 수 있다. 상기 마이크(220)는 항상 활성화되는 제1 마이크(221)(예: 저전력 마이크)와, 프로세서(260)의 명령어에 따라 활성화 또는 비활성화되는 제2 마이크(222)(예: 고성능 마이크)를 포함할 수 있다. 상기 제2 마이크(222)는 적어도 지능형 앱 또는 인텔리전트 어시스턴스 서비스가 실행되지 않은 상태에서는 비활성화될 수 있다. 본 문서에서는 마이크(220)가 제1 마이크(221)와 제2 마이크(222)를 포함하는 경우를 예로 들어 설명한다. 하지만, 이에 한정되지 않는다. 예를 들어, 마이크(220)는 항상 활성화되는 제1 마이크(221)만을 포함하거나, 세 개 이상의 마이크들(예: 제1 마이크(221)와 제2 마이크(222)를 포함함)을 포함할 수 있다.

[0020] 일 실시예에 따르면, 통신 회로(230)는 지정된 통신 채널을 통해 지능형 서버(10)와 통신할 수 있다. 상기 지정된 통신 채널은 예를 들면, 3G, 4G, 5G, Wifi 등의 무선 통신 방식의 통신 채널을 포함할 수 있다.

[0021] 일 실시예에 따르면, 스피커(240)는 프로세서(260)의 명령어에 따른 소리를 출력할 수 있다.

- [0022] 일 실시예에 따르면, 메모리(250)는, 예를 들면, 전자 장치(20)의 적어도 하나의 다른 구성요소에 관계된 명령 또는 데이터를 저장할 수 있다. 메모리(250)는 실행될 때에 상기 프로세서(260)로 하여금, 상기 마이크(220)(예: 제1 마이크(221))를 통해 웨이크업 발화(예: 하이 빅스비)를 수신하면, 음성 인식 서비스를 제공하는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행하고, 상기 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 상기 마이크(220)를 통해 수신된 제1 사용자 발화에 기반하여 결정된 액션을 수행하고, 상기 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 상기 제1 액션을 수행 완료한 시점으로부터 선택된 시간(선택된 시간) 내에 상기 마이크(220)를 통해 수신된 제2 사용자 발화에 기반하여 트리거 단어가 인식되는지를 확인하고, 상기 제2 사용자 발화에 기반하여 상기 트리거 단어가 인식되면, 제2 사용자 발화에 기반하여 결정된 제2 액션을 수행하고, 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제2 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료하도록 설정된 인스트럭션들(instructions)을 저장할 수 있다. 상기 웨이크업 발화는 특정 단어 또는 구문을 포함하고, 특정 사용자의 음향 특성(예: 특정 주파수 특성)을 갖는 발화일 수 있다. 상기 선택된 시간은 상기 제1 액션을 수행 완료한 이후부터의 시간을 포함할 수 있다. 상기 트리거 단어는 입력 회로(210)를 통한 사용자 입력에 의하여 선택되거나, 프로세서(260)에 의하여 선택된 단어 또는 구문을 포함할 수 있다.
- [0023] 일 실시예에 따르면, 프로세서(260)는 메모리(250)에 저장된 인스트럭션들을 이용하여 전자 장치(20)의 적어도 하나의 다른 구성요소들의 제어 및/또는 통신에 관한 연산이나 데이터 처리를 실행할 수 있다. 프로세서(260)는 제1 마이크(221)를 통해 웨이크업 발화를 수신하면(대기 모드), 음성 인식 서비스를 제공하는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다. 이와 달리, 프로세서(260)는 입력 회로(예: 터치 센서 또는 물리적 버튼)를 통해 지정된 입력(예: 터치 또는 눌림)을 확인하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다.
- [0024] 일 실시예에 따르면, 프로세서(260)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 제2 마이크(222)를 통해 수신된 제1 사용자 발화에 기반하여 결정된 액션을 수행할 수 있다. 예를 들어, 프로세서(260)는 제1 사용자 발화를 수신하면, 제1 사용자 발화를 제1 음성 데이터로 변환하고 변환된 제1 음성 데이터를 통신 회로(230)를 통해 지능형 서버(10)로 송신할 수 있다(**음성 인식 모드**). 지능형 서버(10)는 제1 음성 데이터를 수신하면 수신된 제1 음성 데이터를 텍스트로 변환하고, 변환된 텍스트에 기반하여 사용자의 의도를 결정할 수 있다. 지능형 서버(10)는 결정된 사용자의 의도에 따른 제1 액션을 결정하고 결정된 제1 액션 수행과 관련된 정보를 생성하고 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다(**이해 모드**). 전자 장치(20)는 제1 액션 수행과 관련된 정보를 수신하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 제1 액션 수행과 관련된 정보에 기반하여 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행할 수 있다(**액션 수행 모드**). 프로세서(260)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 결정된 액션의 수행 과정 또는 수행 완료 시에 스피커(240)를 통해 액션 수행과 관련된 제1 소리를 출력할 수 있다. 상기 액션 수행과 관련된 제1 소리는 예를 들면, 지능형 서버(10)로부터 제공된 소리일 수 있고, 또는 전자 장치(20)가 액션 수행과 관련된 정보를 기반으로 생성한 소리일 수 있다. 상기 액션 수행과 관련된 제1 소리는 예를 들면, 액션 수행 과정을 안내하는 소리 또는 액션의 수행 결과를 안내하는 소리일 수 있다.
- [0025] 일 실시예에 따르면, 프로세서(260)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제2 마이크(222)를 통해 수신된 제2 사용자 발화에 기반하여 트리거 단어가 인식되는지를 확인할 수 있다(**대화 모드**). 예를 들어, 프로세서(260)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 제2 사용자 발화를 제2 음성 데이터로 변환하고 제2 음성 데이터를 텍스트로 변환하고 변환된 텍스트에 트리거 단어가 포함되는지를 확인할 수 있다. 상기 선택된 시간(예: 약 30분)은 디폴트로 설정된 시간 또는 사용자에게 의해 변경 설정된 시간일 수 있다. 상기 트리거 단어는 예를 들면, 상기 전자 장치(20)가 수행 가능한 복수의 액션들 중에서 상기 입력 회로(210)를 통해 지정된 액션 요청과 관련된 단어를 포함할 수 있다. 상기 트리거 단어는 다른 예를 들면, 전자 장치(20)에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로(210)를 통해 지정된 카테고리에 속하는 액션 요청과 관련된 단어(카테고리에 관련된 단어)를 포함할 수 있다. 상기 트리거 단어는 상기 전자 장치에서 수행 가능한 복수의 액션들 중에서 하나의 액션을 요청하는 구와, 지정된 화제 전환 단어 또는 상기 전자 장치(20)를 지칭하는 단어 중 적어도 하나를 포함할 수 있다.
- [0026] 일 실시예에 따르면, 프로세서(260)는 제2 사용자 발화에 기반하여 트리거 단어를 인식하기에 앞서, 제2 사용자 발화에 기반하여 발화 속도 변화, 톤 변화 및 억양 변화를 확인할 수 있다. 프로세서(260)는 확인된 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 제1 지정된 조건을 만족하는지를 확인하고, 확인된 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 제1 지정된 조건을 만족하는 경우, 제2 사용자 발화에 기반하여 상기 트리거

단어가 인식되는지를 확인할 수 있다. 예를 들어, 프로세서(260)는 발화 속도가 지정된 속도 이상 감소되고, 톤(주파수)이 지정된 주파수 이하로 낮아지고 억양의 변화가 지정된 범위 내이면, 확인된 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 제1 지정된 조건을 만족하는 것으로 판단할 수 있다. 프로세서(260)는 확인된 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 제1 지정된 조건을 만족하지 않는 경우, 제2 사용자 발화에 기반하여 트리거 단어를 인식하지 않을 수 있다. 예를 들어, 프로세서(260)는 확인된 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 제1 지정된 조건을 만족하지 않는 경우, 제2 사용자 발화를 제2 음성 데이터로 변환하지 않을 수 있다.

[0027] 일 실시예에 따르면, 프로세서(260)는 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 트리거 단어가 인식되면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 트리거 단어를 포함하는 문장(텍스트)을 결정할 수 있다. 예를 들어, 프로세서(260)는 제2 마이크(222)를 통해 수신된 제2 사용자 발화를 텍스트로 변환하는 과정에서, 제2 사용자 발화의 수신 시간 간격(예: 제2 마이크(222)의 출력 신호 간격)을 확인하고, 확인된 시간 간격에 기반하여 제2 사용자 발화의 종점(end point)을 확인함에 따라 트리거 단어를 포함하는 문장을 결정할 수 있다.

[0028] 프로세서(260)는 트리거 단어를 포함하는 문장을 통신 회로(230)를 통해 지능형 서버(10)로 송신할 수 있다(이해 모드). 지능형 서버(10)는 트리거 단어를 포함하는 문장을 수신하면, 수신된 문장에 기반하여 사용자의 의도를 결정할 수 있다. 지능형 서버(10)는 결정된 사용자의 의도에 따른 제2 액션을 결정하고 결정된 제2 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 제2 액션 수행과 관련된 정보를 수신하면, 제2 액션 수행과 관련된 정보에 기반하여 제2 사용자 발화에 기반하여 결정된 제2 액션을 수행할 수 있다. 프로세서(260)는 결정된 제2 액션의 수행 과정 또는 수행 완료 시에 스피커(240)를 통해 액션 수행과 관련된 제2 소리를 출력할 수 있다.

[0029] 일 실시예에 따르면, 프로세서(260)는 트리거 단어를 포함하는 문장을 결정하고 결정된 문장을 지능형 서버(10)로 송신하기 전에, 결정된 문장에서 트리거 단어의 위치에 기반하여 상기 문장의 오인식 여부를 결정할 수 있다(대화 모드). 예를 들어, 프로세서(260)는 트리거 단어가 상기 문장의 끝 부분에 위치하는 경우, 상기 문장이 오인식되지 않았다고 결정할 수 있다. 다른 예를 들어, 프로세서(260)는 트리거 단어가 상기 문장의 앞 부분에 위치하는 경우, 상기 문장이 오인식되었다고 판단할 수 있다. 프로세서(260)는 트리거 단어를 포함하는 문장이 오인식된 것으로 결정한 경우, 트리거 단어를 포함하는 문장을 지능형 서버(10)로 송신하지 않을 수 있다. 추가적으로, 프로세서(260)는 트리거 단어를 포함하는 문장이 오인식된 것으로 결정한 경우, 트리거 단어를 포함하는 문장에 대한 재발화를 요청하는 소리를 스피커(240)를 통해 출력할 수 있다. 예를 들어, 프로세서(260)는 트리거 단어 “알려줘”를 인식하고 알려주어야 하는 대상을 확인하지 못한 경우에, “무엇을 알려달라는 거죠”와 같은 재발화를 요청하는 소리를 출력할 수 있다.

[0030] 일 실시예에 따르면, 프로세서(260)는 트리거 단어를 포함하는 문장을 결정하고 결정된 문장을 지능형 서버(10)로 송신하기 전에, 트리거 단어를 포함하는 문장에 리젝(reject) 단어 - 다른 선택된 단어 또는 구문 - 가 포함되는지를 확인하고 트리거 단어를 포함하는 문장에 리젝 단어가 포함되면, 지능형 서버(10)로 상기 문장을 송신하지 않을 수 있다. 상기 리젝 단어는 예를 들면, 입력 회로(210)를 통해 사용자에게 의해 선택되거나, 프로세서(260)의 동작에 의해 선택된 다른 단어 또는 구문을 포함할 수 있다. 다양한 실시 예에 따르면, 프로세서(260)는 제2 사용자 발화에 기반하여 인식된 각 문장에 리젝 단어가 포함되는지를 먼저 확인하고, 리젝 단어가 포함되지 않은 경우에만, 제2 사용자 발화에 트리거 단어가 포함되는지를 확인할 수 있다.

[0031] 일 실시예에 따르면, 프로세서(260)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 이용하여, 제2 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제3 사용자 발화에 기반하여 트리거 단어가 인식되는지를 확인할 수 있다. 프로세서(260)는 제2 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제3 사용자 발화에 기반하여 트리거 단어가 인식되면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스의 실행 시간을 연장하고, 제2 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제3 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다.

[0032] 일 실시예에 따르면, 프로세서(260)는 제1 액션을 수행 완료한 시점으로부터 선택된 시간 내에 제2 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다(대기 모드). 지능형 앱 또는 인텔리전트 어시스턴스 서비스가 종료되면, 프로세서(260)는 웨이크업 발화 또는 지정된 입력을 수신한 이후에 음성 인식 서비스를 제공할 수 있다.

[0033] 다양한 실시 예에 따르면, 프로세서(260)는 제2 사용자 발화에 기반하여 음성 인식 서비스의 종료를 요청하는

중단 요청 단어 - 또 다른 선택된 단어 또는 구문 - 를 인식하면, 선택된 시간이 만료되지 않더라도 제2 마이크 (222)를 비활성화하고 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다.

- [0034] 다양한 실시예에 따르면, 프로세서(260)는 입력 회로(210)를 통한 대화 모드의 사용 설정 여부를 확인하고, 대화 모드가 사용 설정되었는지 여부에 따라 제1 사용자 발화에 기반하여 제1 액션을 수행 완료한 이후에 지능형 앱 또는 인텔리전트 어시스턴스 서비스의 실행 시간을 연장할지 아니면, 종료할지를 결정할 수 있다. 예를 들어, 프로세서(260)는 입력 회로(210)를 통해 대화 모드(또는, 대화 기능)의 사용이 설정된 경우에는 제1 액션을 수행한 후 적어도 한번 제2 사용자 발화에 기반하여 선택된 시간 내에 트리거 단어가 인식되는지를 모니터링할 수 있다. 반면, 프로세서(260)는 입력 회로(210)를 통해 대화 모드가 사용 설정되지 않은 경우에는 제1 액션을 수행한 후 바로 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다. 다양한 실시예에 따르면, 대화 모드의 사용 설정은 마이크(220)를 통해 수신된 사용자 발화에 기반하여 설정될 수 있다.
- [0035] 다양한 실시예에 따르면, 프로세서(260)는 트리거 단어를 포함하는 문장과 더불어, 트리거 단어를 포함하는 문장의 이전 문장(적어도 하나의 문장을 포함할 수 있음)을 지능형 서버(10)로 송신할 수 있다.
- [0036] 다양한 실시예에 따르면, 전자 장치(20)는 제1 사용자 발화 또는 제2 사용자 발화에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 액션 수행에 관련된 정보를 생성 가능한 경우(예: NLU 모듈을 포함하는 경우), 지능형 서버(10)와 통신하지 않고 자체적으로 제1 사용자 발화 또는 제2 사용자 발화에 기반하여 사용자의 의도를 결정하고, 결정된 의도에 따른 액션 수행과 관련된 정보를 생성할 수 있다.
- [0037] 상술한 실시예에 따르면, 전자 장치(20)는 웨이크업 발화(또는, 지정된 입력)를 수신하고 음성 인식 서비스를 시작하면, 사용자 발화들에 트리거 단어가 포함되어 있는지를 실시간으로 모니터링하고, 웨이크업 발화를 대체하는 사용자 발화들의 일부(트리거 단어)에 기반하여 사용자 발화에 대응하는 액션을 수행할 수 있다. 따라서, 사용자가 음성 인식 서비스를 이용하는 복수의 액션들을 요청하기 위하여, 웨이크업 발화를 반복적으로 수행하여야 하던 종래의 불편함을 해소할 수 있다.
- [0038] 일 실시예에 따르면, 시스템(예: 도 1의 통합 지능 시스템(10, 20))은, 마이크(예: 도 2의 마이크(220)); 스피커(예: 도 2의 스피커(240)); 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서(예: 도 2의 프로세서(260)); 및 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리(예: 도 2의 메모리(250))를 포함할 수 있다. 상기 메모리는, 자동 음성 인식 (ASR) 모듈(예: 도 6의 자동 음성 인식 모듈(620)) 및 자연어 이해 (NLU) 모듈(예: 도 6의 NLU 모듈(640))을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 상기 마이크를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 웨이크업(wake-up) 발화(utterance)를 수신하고, 상기 웨이크업 발화 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고, 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고, 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고, 상기 웨이크업 발화 수신 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고, 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고, 상기 제2 사용자 발화에 대한 텍스트에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상기 제2 사용자 발화에 대한 텍스트에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장할 수 있다.
- [0039] 상기 인스트럭션들은, 상기 프로세서가, 상기 선택된 단어 또는 구문을 수신하도록 구성된 사용자 인터페이스를 제공하거나, 상기 프로세서의 동작에 의해 상기 선택된 단어 또는 구문을 결정하도록 할 수 있다.
- [0040] 상기 인스트럭션들은, 상기 프로세서가, 상기 선택된 시간 구간을 수신하도록 구성된 사용자 인터페이스를 제공할도록 할 수 있다.
- [0041] 상기 선택된 시간은, 상기 제1 응답이 생성된 이후부터의 시간을 포함할 수 있다.
- [0042] 상기 선택된 단어 또는 구문은, 카테고리에 관련된 단어를 포함할 수 있다.
- [0043] 상기 인스트럭션들은, 상기 프로세서가, 상기 제1 사용자 발화에 기반하여 발화 속도 변화, 톤 변화 및 억양 변화를 확인하고, 상기 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 지정된 조건을 만족하는 경우, 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 포함되는지를 확인하도록 할 수 있다.
- [0044] 상기 인스트럭션들은, 상기 프로세서가, 상기 제2 사용자 발화에, 상기 선택된 단어 또는 구문과 상이한 다른 선택된 단어 또는 구문이 포함되는지를 확인하고, 상기 제2 사용자 발화에, 상기 다른 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상

기 제2 사용자 발화에, 상기 다른 선택된 단어 또는 구문이 포함되어 있으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 할 수 있다.

- [0045] 상기 인스트럭션들은, 상기 프로세서가, 상기 선택된 단어 또는 구문을 포함하는 문장에서 상기 선택된 단어 또는 구문의 위치에 기반하여, 상기 문장의 오인식 여부를 확인하고, 상기 문장이 오인식된 경우, 상기 선택된 단어 또는 구문을 포함하는 문장에 대한 재발화를 요청하는 소리를 상기 스피커를 통해 출력하도록 할 수 있다.
- [0046] 일 실시예에 따르면, 시스템(예: 도 1의 통합 지능 시스템(10, 20))은, 사용자 인터페이스(예: 도 2의 입력 회로(210)); 마이크(예: 도 2의 마이크(220)); 스피커(예: 도 2의 스피커(240)); 상기 사용자 인터페이스, 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서(예: 도 2의 프로세서(260)); 및 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리(예: 도 2의 메모리(250))를 포함할 수 있다. 상기 메모리는, 자동 음성 인식 (ASR) 모듈 및 자연어 이해 (NLU) 모듈을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 상기 사용자 인터페이스를 통해, 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 사용자 입력을 수신하고, 상기 사용자 입력을 수신한 이후에 상기 마이크를 통해, 제1 사용자 발화를 수신하고, 상기 NLU 모듈을 이용하여 상기 제1 사용자 발화를 처리하고, 상기 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성하고, 상기 사용자 입력을 수신한 후 선택된 시간 구간 동안, 상기 마이크를 통해, 제2 사용자 발화를 수신하고, 상기 ASR 모듈을 이용하여 상기 제2 사용자 발화에 대한 텍스트를 추출하고, 상기 제2 사용자 발화에, 선택된 단어 또는 구문이 포함되어 있으면, 상기 NLU 모듈을 이용하여, 상기 제2 사용자 발화를 처리하여 제2 응답을 생성하고, 상기 제2 사용자 발화에, 상기 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 제2 사용자 발화에 대한 응답을 생성하지 않도록 하는 인스트럭션들을 저장할 수 있다.
- [0047] 상기 사용자 인터페이스는, 터치스크린 디스플레이, 프레스 가능한 버튼, 또는 터치 가능한 버튼 중 적어도 하나를 포함할 수 있다.
- [0048] 상기 시스템은 상기 선택된 단어 또는 구문을 수신하도록 구성된 사용자 인터페이스를 제공하거나, 상기 프로세서의 동작에 의해 상기 선택된 단어 또는 구문을 결정하도록 할 수 있다. 일 실시예에 따르면, 전자 장치(예: 도 2의 전자 장치(20))는 통신 회로(예: 통신 회로(230)); 마이크(예: 도 2의 마이크(220)); 및 상기 통신 회로, 상기 입력 회로 및 상기 마이크와 기능적으로 연결된 프로세서(예: 도 2의 프로세서(260)); 및 상기 프로세서와 기능적으로 연결된 메모리(예: 도 2의 메모리(250))를 포함할 수 있다. 상기 메모리는, 실행될 때에 상기 프로세서로 하여금, 상기 마이크를 통해 음성 인식 서비스를 호출하는 웨이크업 발화를 수신하면, 상기 음성 인식 서비스를 제공할 수 있는 지능형 앱을 실행하고, 상기 지능형 앱을 이용하여, 상기 마이크를 통해 제1 사용자 발화를 수신하고, 상기 지능형 앱을 이용하여, 상기 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행하고, 상기 제1 액션을 수행한 시점으로부터 선택된 시간 내에 상기 마이크를 통해 제2 사용자 발화를 수신하고 상기 지능형 앱을 이용하여, 상기 제2 사용자 발화에 기반하여 선택된 단어 또는 구문이 인식되는지를 확인하고, 상기 선택된 시간 내에 상기 제2 사용자 발화에 기반하여 상기 선택된 단어 또는 구문이 인식되면, 상기 지능형 앱을 이용하여, 상기 제2 사용자 발화를 기반하여 결정된 제2 액션을 수행하고, 상기 선택된 시간 내에 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 인식되지 않으면, 상기 지능형 앱을 종료하도록 설정된 인스트럭션들을 저장할 수 있다.
- [0049] 상기 인스트럭션들은, 상기 프로세서가, 상기 제1 액션을 수행한 시점으로부터 상기 선택된 시간 내에 상기 선택된 단어 또는 구문이 인식되면, 상기 선택된 단어 또는 구문을 포함하는 문장을 결정하고, 상기 문장을 상기 통신 회로를 통해 외부 전자 장치로 송신하고, 상기 외부 전자 장치로부터 상기 문장에 기반하여 결정된 제2 액션 수행과 관련된 정보를 수신하고, 상기 제2 액션 수행과 관련된 정보에 기반하여 상기 제2 액션을 수행하도록 더 설정될 수 있다.
- [0050] 상기 인스트럭션들은, 상기 프로세서가, 상기 문장에 다른 선택된 단어 또는 구문이 포함되는지를 확인하고, 상기 문장에 상기 다른 선택된 단어 또는 구문이 포함되어 있지 않으면, 상기 문장을 상기 통신 회로를 통해 상기 외부 전자 장치로 송신하고, 상기 문장에 상기 다른 선택된 단어 또는 구문이 포함되어 있으면, 상기 문장을 상기 외부 전자 장치로 송신하지 않도록 더 설정될 수 있다.
- [0051] 상기 전자 장치는 스피커(예: 도 2의 스피커(240))를 더 포함하고, 상기 인스트럭션들은, 상기 프로세서가, 상기 문장에서 상기 선택된 단어 또는 구문의 위치에 기반하여, 상기 확인된 문장의 오인식 여부를 확인하고, 상기 문장의 오인식된 경우, 상기 선택된 단어 또는 구문을 포함하는 문장에 대한 재발화를 요청하는 소리를 상기 스피커를 통해 출력하도록 더 설정될 수 있다.

- [0052] 상기 인스트럭션들은, 상기 프로세서가, 상기 제2 사용자 발화에 기반하여 또 다른 선택된 단어 또는 구문이 인식되면, 상기 지능형 앱을 종료하도록 더 설정될 수 있다.
- [0053] 상기 선택된 단어 또는 구문은, 상기 전자 장치에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로를 통해 지정된 액션 요청과 관련된 단어를 포함할 수 있다.
- [0054] 상기 선택된 단어 또는 구문은, 상기 전자 장치에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로를 통해 지정된 카테고리에 속하는 액션 요청과 관련된 단어를 포함할 수 있다.
- [0055] 상기 선택된 단어 또는 구문은, 상기 전자 장치에서 수행 가능한 복수의 액션들을 요청하는 단어와, 화제 전환을 위한 단어 및 상기 전자 장치를 지칭하는 단어 중 적어도 하나를 더 포함할 수 있다.
- [0056] 상기 인스트럭션들은, 상기 프로세서가, 상기 제1 사용자 발화에 기반하여 발화 속도 변화, 톤 변화 및 억양 변화를 확인하고, 상기 발화 속도 변화, 상기 톤 변화 및 상기 억양 변화가 지정된 조건을 만족하는 경우, 상기 제2 사용자 발화에 상기 선택된 단어 또는 구문이 포함되는지를 확인하도록 더 설정될 수 있다.
- [0057] 도 3a는 일 실시예에 따른 대화 모드가 사용 설정되지 않은 경우의 전자 장치의 음성 인식 서비스를 위한 동작 모드들을 나타낸다.
- [0058] 도 3a를 참조하면, 대기 모드(310)에서, 전자 장치(20)는 마이크(예: 제1 마이크(221))를 통해 웨이크업 발화(예: 하이 빅스비)를 수신하거나, 입력 회로(210)를 통해 지정된 입력을 수신하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행하여 음성 수신 모드(320)로 전환할 수 있다. 상기 대기 모드(310)에서, 제1 마이크(221)는 활성화되고, 제2 마이크(222)는 비활성화되고, 지능형 앱 또는 인텔리전트 어시스턴스 서비스는 실행되지 않을 수 있다. 음성 수신 모드(320)에서, 전자 장치(20)는 제2 마이크(222)를 활성화하고, 음성 인식 서비스를 제공하는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다.
- [0059] 음성 수신 모드(320)에서, 전자 장치(20)는 제2 마이크(222)를 통해 수신된 제1 사용자 발화를 제1 음성 데이터로 변환하고 상기 제1 음성 데이터를 통신 회로(230)를 통해 지능형 서버(10)로 송신할 수 있다. 이 경우, 지능형 서버(10)는 수신한 음성 데이터를 텍스트로 변환할 수 있다.
- [0060] 전자 장치(20)는 제1 사용자 발화의 수신이 완료되면, 음성 수신 모드(320)에서 이해 모드(330)로 전환할 수 있다. 예를 들어, 전자 장치(20)는 다른 선택된 시간 동안 제1 사용자 발화가 수신되지 않으면, 제1 사용자 발화의 수신이 완료된 것으로 결정하고, 음성 수신 모드(320)에서 이해 모드(330)로 전환할 수 있다. 이해 모드(330)에서, 지능형 서버(10)는 제1 사용자 발화에 대한 텍스트를 기반으로 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제1 액션을 결정할 수 있다. 지능형 서버(10)는 결정된 제1 액션 수행에 관련된 정보를 전자 장치(20)로 송신할 수 있다.
- [0061] 전자 장치(20)는 지능형 서버(10)로부터 제1 액션 수행과 관련된 정보를 수신하면, 액션 수행 모드(340)로 전환할 수 있다. 액션 수행 모드(340)에서, 전자 장치(20)는 지능형 서버(10)로부터 수신된 제1 액션 수행과 관련된 정보에 기반하여 제1 사용자 발화에 대응하는 제1 액션을 수행할 수 있다. 전자 장치(20)는 제1 액션을 수행 완료하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료하고 대기 모드(310)로 전환할 수 있다.
- [0063] 도 3b는 일 실시예에 따른 대화 모드가 사용 설정되지 않은 경우의 음성 인식 서비스의 예를 나타낸다.
- [0064] 도 3b를 참조하면, 대화 모드가 사용 설정되지 않은 경우, 전자 장치(20)는 웨이크업 발화(3511)을 수신하면(대기 모드), 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다(음성 인식 모드). 전자 장치(20)는 음성 인식 모드에서 사용자 발화(352) “내일 날씨 알려줘”를 수신하고, 사용자 발화(352)를 제1 음성 데이터로 변환하고, 제1 음성 데이터를 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 제1 음성 데이터에 기반하여 제1 사용자 발화에 대응하는 제1 액션을 결정하고, 제1 액션 수행과 관련한 정보를 전자 장치(20)로 송신할 수 있다(이해 모드). 전자 장치(20)는 지능형 서버(10)로부터 제1 액션 수행과 관련된 정보를 수신하고, 제1 액션 수행과 관련된 정보에 기반하여 금요일 날씨를 사용자에게 알려주는 제1 액션(353) 예컨대, “금요일 날씨는 화창하겠습니다” 소리를 출력하는 제1 액션을 수행할 수 있다(액션 수행 모드). 전자 장치(20)는 제1 액션을 수행한 후 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료하고 다시 대기 모드로 전환할 수 있다.
- [0065] 대기 모드에서, 전자 장치(20)는 다시 웨이크업 발화(354)를 수신하고, 음성 수신 모드로 전환하여 사용자 발화(355) “수원 맛집 알려줘”를 수신하여 지능형 서버(10)로 송신할 수 있다. 전자 장치(20)는 사용자 발화(35

5)을 수신 완료하면, 이해 모드로 전환할 수 있다. 이해 모드에서 지능형 서버(10)는 사용자 발화(355)에 기반하여 제2 액션을 결정하고, 결정된 제2 액션 수행과 관련한 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 지능형 서버(10)로부터 사용자 발화(355)에 기반하여 결정된 제2 액션 수행과 관련된 정보를 수신하면, 이해 모드에서 액션 수행 모드로 전환할 수 있다. 액션 수행 모드에서 전자 장치(20)는 수신된 액션 수행과 관련된 정보에 기반하여 수원 맛집을 사용자에게 알려주는 액션(356) 예컨대, “수원 맛집은 XXX와 XXA가 있습니다” 소리를 출력하는 액션을 수행한 후 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료하고 다시 대기 모드로 전환할 수 있다.

[0066] 대기 모드에서 전자 장치(20)는 다시 웨이크업 발화(357)를 수신하여 음성 수신 모드로 전환하여 사용자 발화(358) “토요일 날씨 알려줘” 를 수신하여 지능형 서버(10)로 송신할 수 있다. 전자 장치(20)는 사용자 발화(358)을 수신 완료하면, 이해 모드로 전환할 수 있다. 이해 모드에서 지능형 서버(10)는 사용자 발화(358)에 기반하여 제3 액션을 결정하고, 제3 액션 수행과 관련한 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 지능형 서버(10)로부터 사용자 발화(358)에 기반하여 결정된 제3 액션 수행과 관련된 정보를 수신하면, 이해 모드에서 액션 수행 모드로 전환할 수 있다. 액션 수행 모드에서 전자 장치(20)는 수신된 액션 수행과 관련된 정보에 기반하여 토요일 날씨를 사용자에게 알려주는 액션(359) 예컨대, “토요일 날씨는 흐리겠습니다” 소리를 출력하는 액션을 수행하고 다시 대기 모드로 전환할 수 있다.

[0067] 상술한 실시예와 같이, 전자 장치(20)는 입력 회로(210)를 통해 대화 모드가 사용 설정되지 않은 경우, 세 가지 사용자 발화(352, 355, 358)에 대응하는 3가지의 액션(353, 356, 359)을 수행하기 위해서 총 3번의 웨이크업 발화(351, 354, 357)를 필요로 하여 사용자의 번거로움을 야기할 뿐 아니라, 음성 인식 서비스를 위한 지능형 앱 또는 인텔리전트 어시스턴스 서비스의 실행과 실행 종료를 반복하여 처리 부하가 크고 프로세싱이 복잡할 수 있다.

[0069] 도 4a는 일 실시예에 따른 대화 모드가 사용 설정된 경우의 전자 장치의 음성 인식 서비스를 위한 동작 모드들을 나타낸다.

[0070] 도 4a를 참조하면, 대화 모드가 사용 설정된 경우에, 전자 장치(20)는 대기 모드에서 제1 마이크(221)를 통해 웨이크업 발화를 인식하거나, 입력 회로(210)를 통해 지정된 입력을 수신하면, 음성 수신 모드로 전환할 수 있다. 전자 장치(20)는 지능형 앱 또는 인텔리전트 어시스턴스 서비스에 의해 음성 수신 모드에서 제2 마이크(222)를 활성화하고 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다. 전자 장치(20)는 음성 수신 모드에서 제1 사용자 발화를 수신 완료하면, 이해 모드로 전환할 수 있다. 지능형 서버(10)는 제1 사용자 발화에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제1 액션을 결정할 수 있다. 지능형 서버(10)는 결정된 제1 액션 수행과 관련된 정보를 생성하고, 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 제1 액션 수행과 관련된 정보를 수신하면, 이해 모드에서 액션 수행 모드로 전환할 수 있다. 전자 장치(20)는 액션 수행 모드에서, 지능형 앱 또는 인텔리전트 어시스턴스 서비스에 의해, 제1 액션 수행과 관련된 정보에 기반하여 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행할 수 있다.

[0071] 전자 장치(20)는 대화 모드가 사용 설정된 경우, 액션 수행 모드에서 대화 모드로 전환할 수 있다. 전자 장치(20)는 대화 모드에서 액션 수행 완료 후 선택된 시간 내에 제2 마이크(222)를 통해 수신된 제2 사용자 발화에 기반하여 트리거 단어가 인식되는지를 확인할 수 있다. 예를 들어, 전자 장치(20)는 제2 마이크(222)를 통해 수신된 제2 사용자 발화를 텍스트로 변환하고 변환된 텍스트로부터 트리거 단어가 인식되는지를 확인할 수 있다. 전자 장치(20)는 제2 사용자 발화에 기반하여 트리거 단어가 인식되면, 트리거 단어를 포함하는 문장을 결정하고, 결정된 문장을 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 전자 장치(20)로부터 수신된 문장에 기반하여 제2 사용자 발화에 기반하여 사용자의 의도를 결정하고, 결정된 사용자 의도에 따른 제2 액션을 결정할 수 있다. 지능형 서버(10)는 수신된 문장에 기반하여 결정된 제2 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 제2 액션 수행과 관련된 정보를 수신하면, 이해 모드에서 액션 수행 모드로 전환할 수 있다. 전자 장치(20)는 액션 수행 모드에서 제2 액션 수행과 관련된 정보에 기반하여 토요일 날씨를 사용자에게 알려주는 제2 액션(390)을 수행할 수 있다. 전자 장치(20)는 제2 액션 수행을 완료하면, 액션 수행 모드에서 다시 대화 모드로 전환할 수 있다.

[0072] 전자 장치(20)는 대화 모드에서 제2 마이크(222)를 통해 수신된 제2 사용자 발화에 기반하여 트리거 단어를 인식하지 못한 경우, 대기 모드로 전환할 수 있다. 전자 장치(20)는 대기 모드로 전환되기 전에 제2 마이크(222)

를 비활성화하고 지능형 앱 또는 인텔리전트 어시스턴스 서비스의 실행을 종료할 수 있다.

- [0074] 도 4b는 일 실시예에 따른 대화 모드가 사용 설정된 경우의 음성 인식 서비스의 예를 나타낸다.
- [0075] 도 4b를 참조하면, 전자 장치(20)는 대화 모드가 사용 설정된 경우, 전자 장치(20)는 대기 모드에서 웨이크업 발화(410)을 수신하면 음성 수신 모드로 전환하여 수신된 제1 사용자 발화(420) “내일 날씨 알려줘”를 지능형 서버(10)로 송신할 수 있다. 전자 장치(20)는 제1 사용자 발화(420)을 수신 완료하면, 이해 모드로 전환할 수 있다. 지능형 서버(10)는 이해 모드에서 제1 사용자 발화(420)에 기반하여 결정된 액션 수행과 관련한 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 지능형 서버(10)로부터 제1 사용자 발화(420)에 기반하여 결정된 제1 액션 수행과 관련된 정보를 수신하면, 이해 모드로부터 액션 수행 모드로 전환하고, 액션 수행 모드에서 수신된 제1 액션 수행과 관련된 정보에 기반하여 “금요일 날씨는 화창하겠습니다”라는 금요일 날씨를 안내하는 제1 소리를 출력하는 제1 액션을 수행하고 대화 모드로 전환할 수 있다.
- [0076] 전자 장치(20)는 대화 모드에서 선택된 시간 내에 수신된 사용자 발화(450)에 기반하여 사용자 발화(450)의 일부인 트리거 단어(예: 알려줘)를 인식할 수 있다. 전자 장치(20)는 트리거 단어 “알려줘”를 인식하면, 트리거 단어 “알려줘”를 포함하는 문장 “수원 맛집 알려줘”를 결정하고, 결정된 문장을 지능형 서버(10)로 송신하고, 이해 모드로 전환할 수 있다. 지능형 서버(10)는 전자 장치(20)로부터 트리거 단어를 포함하는 문장 “수원 맛집 알려줘”를 수신하고 “수원 맛집 알려줘”에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 수원 맛집을 알려주는 제2 액션을 결정할 수 있다. 지능형 서버(10)는 수원 맛집을 알려주는 제2 액션 수행과 관련된 정보를 생성하고, 생성된 제2 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다.
- [0077] 전자 장치(20)는 제2 액션 수행과 관련된 정보를 수신하면, 이해 모드에서 액션 수행 모드로 전환할 수 있다. 전자 장치(20)는 액션 수행 모드에서 액션 수행과 관련된 정보에 기반하여 수원 맛집을 안내하는 제2 액션을 수행할 수 있다. 예를 들어, 전자 장치(20)는 주변 맛집을 안내하는 어플리케이션인 망고 플레이트를 실행하고, “망고 플레이트를 수행합니다” 소리를 출력할 수 있다.
- [0078] 상술한 실시예들과 같이, 전자 장치(20)는 입력 회로(210)를 통해 대화 모드가 사용 설정된 경우, 제1 및 제2 사용자 발화(420, 450)에 대응하는 복수의 액션들(430, 460)을 수행하기 위해서 1번의 웨이크업 발화(410)만을 필요로 하므로, 사용자 발화에 기반한 응답을 연속하여 요청하거나, 빈번히 요청하는 경우에 음성 인식 서비스의 편의성을 향상시킬 수 있다.
- [0080] 도 5a는 일 실시예에 따른 대화 모드의 사용을 설정하기 위한 UI 화면을 나타낸다.
- [0081] 도 5a를 참조하면, 화면 510에서, 전자 장치(20)는 예컨대, 입력 회로(210)를 통한 사용자 입력에 따라 위치 이동 가능한, 대화 모드의 사용 설정 또는 비사용 설정을 위한 가상 버튼(511)을 제공할 수 있다. 전자 장치(20)는 가상 버튼(511)이 우측에 위치한 경우, 대화 모드가 사용 설정된 것으로 확인하고 대화 모드에 대한 세부 설정을 위한 가상 버튼(512) 및 가상 버튼(513)을 제공(예: 활성화)할 수 있다. 반면, 전자 장치(20)는 가상 버튼(511)이 좌측에 위치한 경우, 대화 모드가 사용 설정되지 않은 것으로 확인하고, 대화 모드에 대한 세부 설정을 위한 가상 버튼(512) 및 가상 버튼(513)을 제공하지 않을(예: 비활성화)할 수 있다. 다양한 실시 예에 따르면, 화면 510에서, 전자 장치(20)는 대화 모드를 설명하는 문구(515)를 표시할 수 있다. 예를 들어, 전자 장치(20)는 화면 510 중 하단에 “전자 장치가 웨이크업 발화 없이도 선택된 시간 동안 주변 음성을 듣고 대화에 참여할 수 있습니다”라는 대화 모드를 설명하는 문구(515)를 표시할 수 있다.
- [0082] 전자 장치(20)는 대화 모드가 사용 설정된 상태에서, 입력 회로(210)를 통해 가상 버튼(512)이 선택되면, 대화 모드에 대한 시간 설정을 위한 UI 화면을 표시할 수 있다. 다양한 실시 예에 따르면, 전자 장치(20)는 대화 모드가 사용 설정된 상태에서, 입력 회로(210)를 통해 가상 버튼(513)이 선택되면, 전자 장치(20)는 입력 회로(210)를 통해 대화 모드에 대한 트리거 단어를 지정하기 위한 UI 화면을 표시할 수 있다.
- [0084] 도 5b는 일 실시예에 따른 대화 모드에 대한 시간 설정을 위한 UI 화면을 나타낸다.
- [0085] 도 5b를 참조하면, 전자 장치(20)는 화면 520에서, 대화 모드의 시간 설정을 위한 사용자 인터페이스를 표시할 수 있다. 예를 들어, 전자 장치(20)는 도 5a의 화면 510에서 가상 버튼(512)이 선택된 경우에, 화면 520를 표시

할 수 있다. 화면 520은 대화 모드의 선택된 시간을 선택(또는 설정)하기 위한 가상 버튼들(521~524)을 포함할 수 있다. 화면 520에서, 전자 장치(20)는 5분과 관련된 가상 버튼(523)이 선택되면, 선택된 시간을 5분으로 설정할 수 있다. 다른 예를 들어, 대화 모드의 선택된 시간을 임의로 설정하는 기능(화면 520의 사용자 설정)과 관련된 가상 버튼(524)이 선택되면, 사용자 입력에 기반하여 선택된 시간을 임의로 설정 가능한 인터페이스를 제공할 수 있다. 다양한 실시 예에 따르면, 화면 520에서, 전자 장치(20)는 대화 모드에 대한 시간 설정을 설명하는 문구(525)를 표시할 수 있다.

[0087] 도 5c는 일 실시예에 따른 대화 모드에 대한 트리거 단어 설정을 위한 UI 화면의 일 예를 나타낸다.

[0088] 도 5c를 참조하면, 전자 장치(20)는 화면 530에서, 대화 모드에 대한 트리거 단어 설정을 위한 사용자 인터페이스를 표시할 수 있다. 예를 들어, 전자 장치(20)는 도 5a의 화면 510에서 가상 버튼(513)이 선택된 경우에, 화면 530을 표시할 수 있다. 화면 530에서, 전자 장치(20)는 지정 가능한 트리거 단어 목록(531a, 532a, 533a, 534a, 535a) 및 각 트리거 단어들(531a, 532a, 533a, 534a, 535a 중 적어도 하나)을 선택하기 위한 가상 버튼 목록(531b, 532b, 533b, 534b, 535b)을 표시할 수 있다. 화면 530에서, 전자 장치(20)는 입력 회로(210)를 통해 가상 버튼(531b)과 가상 버튼(533b)이 선택되면, 트리거 단어로서 “검색해줘” 및 “알려줘”를 지정할 수 있다. 전자 장치(20)는 화면 530에서, 입력 회로(210)를 통해 트리거 단어를 임의로 설정하는 기능(534a)과 관련된 가상 버튼(534b)이 선택되면, 사용자가 임의로 트리거 단어를 입력 가능한 인터페이스를 제공하고, 제공된 인터페이스를 통해 입력된 단어를 트리거 단어로 설정할 수 있다.

[0089] 다양한 실시예에 따르면, 화면 530에서, 입력 회로(210)를 통해 전자 장치(20)가 자동으로 트리거 단어를 선택하도록 하는 가상 버튼(535b)이 선택되면, 전자 장치(20)(예: 도 2의 프로세서(260))는 지정된 화제 전환 단어(예: 야, 그림) 또는 전자 장치(20)를 지칭하는 단어(예: 빅스비, 빅스) 중 적어도 하나와 전자 장치(20)에서 수행 가능한 복수의 액션들을 각기 요청하는 단어들(예: 검색해줘)를 트리거 단어로 설정(선택)할 수 있다.

[0091] 도 5d는 일 실시예에 따른 대화 모드에 대한 트리거 단어 설정을 위한 UI 화면의 다른 예를 나타낸다.

[0092] 도 5d를 참조하면, 전자 장치(20)는 화면 540에서, 대화 모드에 대한 트리거 단어가 속하는 트리거 카테고리 지정을 위한 사용자 인터페이스를 표시할 수 있다. 예를 들어, 전자 장치(20)는 도 5a의 화면 510에서 가상 버튼(513)이 선택된 경우에, 화면 540을 표시할 수 있다. 화면 540에서, 전자 장치(20)는 설정 가능한 트리거 카테고리 목록(541a, 542a, 543a, 544a, 545a) 및 각 트리거 카테고리들(541a, 542a, 543a, 544a, 545a 중 적어도 하나)을 선택하기 위한 가상 버튼 목록(541b, 542b, 543b, 544b, 545b)을 표시할 수 있다.

[0093] 화면 540에서, 전자 장치(20)는 입력 회로(210)를 통해 가상 버튼(541b) 및 가상 버튼(543b)이 선택되면, 검색 카테고리에 속하는 액션을 요청하는 단어 셋(set)(예: 검색해줘, 찾아줘, 알려줘) 및 뮤직(Music) 카테고리에 속하는 액션을 요청하는 단어(또는, 구문) 셋(set)(예: 음악 틀어줘, 음악 검색해줘, 볼륨 높여줘, 볼륨 낮춰줘)에 포함된 단어들(또는, 구문)을 대화 모드에 대한 트리거 단어로 지정할 수 있다. 이와 관련하여, 전자 장치(20)의 메모리(250)는 각 트리거 카테고리 목록에 속하는 액션을 요청하는 단어 셋(set)을 저장하고, 전자 장치(20)는 입력 회로(210)를 통해 각 트리거 카테고리가 선택되면, 선택된 카테고리에 속하는 액션을 요청하는 단어 셋(set)을 모두 트리거 단어로 지정할 수 있다. 다양한 실시예에 따르면, 화면 540에서, 입력 회로(210)를 통해 전자 장치(20)가 자동으로 트리거 단어를 선택하도록 하는 가상 버튼(545b)이 선택되면, 전자 장치(20)(예: 도 2의 프로세서(260))는 지정된 화제 전환 단어(예: 야, 그림) 또는 전자 장치(20)를 지칭하는 단어(예: 빅스비, 빅스) 중 적어도 하나와 전자 장치(20)에서 수행 가능한 복수의 액션들을 각기 요청하는 단어들(예: 검색해줘)를 트리거 단어로 설정(선택)할 수 있다.

[0095] 도 6은 일 실시예에 따른 프로세서의 구성요소들의 일 예를 나타낸다.

[0096] 도 6을 참조하면, 일 실시예에 따르면, 프로세서(예: 도 2의 프로세서(260))는 클라이언트 모듈(610), 자동 음성 인식 모듈(620) 및 트리거 확인 모듈(630)을 포함할 수 있다. 프로세서(260)의 각 구성요소들은 각각 별도의 하드웨어 모듈이거나 또는 적어도 하나의 프로세서(260)에 의해 구현되는 소프트웨어 모듈일 수 있다. 예를 들어, 프로세서(260)에 포함된 각각의 모듈들이 수행하는 기능은 하나의 프로세서에 의해 수행되거나 또는 각각 별도의 프로세서에 의해 수행될 수 있다.

- [0097] 클라이언트 모듈(610)은 대기 모드에서 웨이크업 발화 이후에 실행되어, 음성 수신 모드에서 제2 마이크(222)를 통해 수신된 제1 사용자 발화를 전처리한 후 제1 음성 데이터로 변환하고, 제1 음성 데이터를 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 제1 음성 데이터를 수신하면, ASR 모듈(미도시)(예: 고성능 ASR 모듈) 및 NLU 모듈(640)을 이용하여 제1 음성 데이터를 기반으로 사용자의 의도를 결정하고 결정된 사용자 의도에 대응하는 제1 액션 수행과 관련된 정보를 생성하고, 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20) 내 지능형 에이전트(270)로 송신할 수 있다(이해 모드). 클라이언트 모듈(610)은 제1 액션 수행과 관련된 정보를 지능형 서버(10)로부터 수신하면, 액션 수행 모드에서 제1 액션 수행과 관련된 정보를 기반으로 제1 액션을 수행할 수 있다.
- [0098] 클라이언트 모듈(610)은 대화 모드에서 제2 마이크(222)를 통해 수신된 사용자 발화를 전처리하여 음성 데이터로 변환할 수 있다. 예를 들어, 클라이언트 모듈(610)은 수신된 사용자 발화로부터 에코와 잡음을 제거하고, 에코와 잡음이 제거된 사용자 발화를 음성 데이터로 변환할 수 있다. 클라이언트 모듈(610)은 수신된 사용자 발화의 시간 간격에 기반하여 종점을 검출하고 종점 정보를 생성할 수 있다. 클라이언트 모듈(610)은 대화 모드에서 변환된 음성 데이터 및 종점 정보를 자동 음성 인식 모듈(620)로 송신할 수 있다.
- [0099] 자동 음성 인식(ASR) 모듈(620)은 대화 모드에서 음성 데이터 및 종점 정보를 수신하면, 제2 음성 데이터를 텍스트로 변환하고, 종점 정보에 기반하여 변환된 텍스트를 문장 단위로 출력할 수 있다.
- [0100] 트리거 확인 모듈(630)은 대화 모드에서 각 문장에 포함된 텍스트와 지정된 트리거 단어(트리거 단어)를 비교하고, 비교 결과에 기반하여 각 문장으로부터 트리거 단어가 인식(예: 포함)되는지를 확인할 수 있다. 트리거 확인 모듈(630)은 선택된 시간 내에 트리거 단어가 인식되면, 트리거 단어를 포함하는 문장을 지능형 서버(10)로 송신할 수 있다. 상기 트리거 단어는 예를 들면, 상기 전자 장치(20) 내 지능형 에이전트(270)에서 수행 가능한 복수의 액션들 중에서 상기 입력 회로(210)를 통해 지정된 액션 요청과 관련된 단어를 포함할 수 있다.
- [0101] 지능형 서버(10)는 이해 모드에서 트리거 단어를 포함하는 문장을 수신하면, NLU 모듈(640)을 이용하여 수신된 문장에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제2 액션을 결정할 수 있다. NLU 모듈(640)은 결정된 제2 액션 수행과 관련된 정보를 생성하고, 생성된 제2 액션 수행과 관련된 정보를 클라이언트 모듈(610)로 송신할 수 있다.
- [0102] 클라이언트 모듈(610) 또는 다른 모듈은 액션 수행 모드에서 제2 액션 수행과 관련된 정보에 기반하여 트리거 단어를 포함하는 문장에 대응하는 제2 액션을 수행할 수 있다.
- [0103] 상술한 실시 예에서, 대화 모드에서(예: 제1 액션의 수행 완료한 시점으로부터) 제2 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 클라이언트 모듈(610), 자동 음성 인식 모듈(620) 및 트리거 확인 모듈(630)의 실행은 종료될 수 있다(대기 모드).
- [0104] 다양한 실시예에 따르면, ASR 모듈(미도시) 및 NLU 모듈(640)은 전자 장치(20) 내 지능형 에이전트(270)에 포함될 수 있다.
- [0106] 도 7는 일 실시예에 따른 프로세서의 구성요소들의 다른 예를 나타낸다.
- [0107] 도 7를 참조하면, 일 실시예에 따르면, 프로세서(260)(예: 도 2의 프로세서(260))는 클라이언트 모듈(710), 자동 음성 인식 모듈(720), 제1 NLU 모듈(730) 및 트리거 확인 모듈(740)을 포함할 수 있다. 프로세서(260)의 각 구성요소들은 각각 별도의 하드웨어 모듈이거나 또는 적어도 하나의 프로세서(260)에 의해 구현되는 소프트웨어 모듈일 수 있다. 예를 들어, 프로세서(260)에 포함된 각각의 모듈들이 수행하는 기능은 하나의 프로세서에 의해 수행되거나 또는 각각 별도의 프로세서에 의해 수행될 수 있다.
- [0108] 클라이언트 모듈(710)은 대기 모드에서 웨이크업 발화 이후에 실행되어, 음성 수신 모드에서 제2 마이크(222)를 통해 수신된 제1 사용자 발화를 전처리한 후 제1 음성 데이터로 변환하고, 제1 음성 데이터를 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 제1 음성 데이터를 수신하면, ASR 모듈(미도시)(예: 고성능 ASR 모듈) 및 NLU 모듈(640)을 이용하여 제1 음성 데이터를 기반으로 사용자의 의도를 결정하고 결정된 사용자 의도에 대응하는 제1 액션 수행과 관련된 정보를 생성하고, 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20) 내 지능형 에이전트(270)로 송신할 수 있다(이해 모드). 클라이언트 모듈(710)은 제1 액션 수행과 관련된 정보를 지능형 서버(10)로부터 수신하면, 액션 수행 모드에서 제1 액션 수행과 관련된 정보를 기반으로 제1 액션을 수행할 수 있다.

- [0109] 클라이언트 모듈(710)은 대화 모드에서 제2 마이크(222)를 통해 수신된 사용자 발화를 전처리하여 음성 데이터로 변환할 수 있다. 예를 들어, 클라이언트 모듈(710)은 수신된 사용자 발화로부터 에코와 잡음을 제거하고, 에코와 잡음이 제거된 사용자 발화를 음성 데이터로 변환할 수 있다. 클라이언트 모듈(710)은 수신된 사용자 발화의 시간 간격에 기반하여 종점을 검출하고 종점 정보를 생성할 수 있다. 클라이언트 모듈(710)은 대화 모드에서 변환된 음성 데이터 및 종점 정보를 자동 음성 인식 모듈(720)로 송신할 수 있다.
- [0110] 자동 음성 인식(ASR) 모듈(720)(예: 저성능 ASR 모듈)은 대화 모드에서 음성 데이터 및 종점 정보를 수신하면, 음성 데이터를 텍스트로 변환하고, 종점 정보에 기반하여 변환된 텍스트를 문장 단위로 출력할 수 있다.
- [0111] 제1 NLU 모듈(730)(예: 저성능 NLU 모듈)은 대화 모드에서 각 문장을 문법적 단위(예: 단어, 구, 형태소)로 나누고, 각 문장에 포함된 단어의 의미를 기반으로 각 문장이 속하는 카테고리를 분류할 수 있다. 예를 들어, 제1 NLU 모듈(730)은 RNN(recurrent neural network) 알고리즘에 기반하여, 메모리(250)에 저장된 제1 자연어 인식 데이터베이스를 이용하여 각 문장에 포함된 단어의 의미가 입력 회로(210)를 통해 선택(또는, 전자 장치(20) 내 지능형 에이전트(270)에 의해 선택)된 카테고리에 속하는지를 확인할 수 있다.
- [0112] 트리거 확인 모듈(740)은 대화 모드에서 각 문장에 포함된 텍스트와, 메모리(250)로부터 획득된, 상기 선택된 카테고리에 관련된 트리거 단어들(트리거 단어)을 비교하고, 비교 결과에 기반하여 각 문장에 적어도 하나의 트리거 단어가 포함(인식)되는지를 확인할 수 있다. 트리거 확인 모듈(740)은 각 문장에 트리거 단어가 포함되면, 트리거 단어를 포함하는 문장을 지능형 서버(10)로 송신할 수 있다. 상기 트리거 단어는 다른 예를 들면, 전자 장치(20) 내 지능형 에이전트(270)가 수행 가능한 복수의 액션들 중에서 상기 입력 회로(210)를 통해 지정된 카테고리에 속하는 액션 요청과 관련된 단어(카테고리에 관련된 단어)를 포함할 수 있다. 또는, 상기 트리거 단어는 상기 전자 장치(20) 내 지능형 에이전트(270)가 수행 가능한 복수의 액션들 중에서 하나의 액션을 요청하는 구와, 지정된 화제 전환 단어 또는 상기 전자 장치(20) 내 지능형 에이전트(270)를 지칭하는 단어 중 적어도 하나를 포함할 수 있다.
- [0113] 지능형 서버(10)는 트리거 단어를 포함하는 문장을 수신하면, 제2 NLU 모듈(750)을 이용하여 수신된 문장에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 액션을 결정할 수 있다. 제2 NLU 모듈(750)은 결정된 액션 수행과 관련된 정보를 생성하고, 생성된 액션 수행과 관련된 정보를 클라이언트 모듈(710)로 송신할 수 있다(이해 모드).
- [0114] 클라이언트 모듈(710) 또는 다른 모듈은 액션 수행과 관련된 정보에 기반하여 트리거 단어를 포함하는 문장에 대응하는 액션을 수행할 수 있다.
- [0115] 상술한 실시 예에서, 대화 모드에서(예: 제1 액션의 수행 완료한 시점으로부터) 제2 사용자 발화에 기반하여 트리거 단어가 인식되지 않으면, 클라이언트 모듈(710), 자동 음성 인식 모듈(720), 제1 NLU 모듈(730) 및 트리거 확인 모듈(740)의 실행은 종료될 수 있다(대기 모드).
- [0116] 다양한 실시예에 따르면, ASR 모듈(미도시) 및 제2 NLU 모듈(750)은 전자 장치(20) 내 지능형 에이전트(270)에 포함될 수 있다.
- [0118] 도 8은 일 실시예에 따른 통합 지능 시스템에 의한 음성 인식 서비스 제공 방법(800)을 나타낸다.
- [0119] 도 8을 참조하면, 동작 810에서, 전자 장치(20)는 마이크(예: 제1 마이크(221))를 통해 음성 기반 인텔리전트 어시스턴스 서비스(음성 인식 서비스)를 호출하는 웨이크업 발화를 수신할 수 있다. 전자 장치(20)는 웨이크업 발화를 수신하면, 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행하고, 제2 마이크(222)를 활성화할 수 있다.
- [0120] 동작 820에서, 전자 장치(20)는 웨이크업 발화 이후에 마이크(예: 제2 마이크(222))를 통해 제1 사용자 발화를 수신할 수 있다.
- [0121] 동작 830에서, 전자 장치(20)는 NLU 모듈(예: 도 7의 제2 NLU 모듈(750))을 이용하여 제1 사용자 발화를 처리할 수 있다. 예를 들어, 전자 장치(20)는 제1 사용자 발화를 제1 음성 데이터로 변환하고 제1 음성 데이터를 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 NLU 모듈(750)을 이용하여, 제1 음성 데이터에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제1 액션을 결정할 수 있다.
- [0122] 동작 840에서, 지능형 서버(10)는 처리된 제1 사용자 발화에 기초하여 제1 응답을 생성할 수 있다. 예를 들어,

지능형 서버(10)는 결정된 제1 액션 수행과 관련된 정보(제1 응답)를 생성하고, 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 지능형 서버(10)로부터 제1 액션 수행과 관련된 정보를 수신하면, 제1 액션 수행과 관련된 정보를 기반으로 제1 사용자 발화에 대응하는 제1 액션을 수행할 수 있다.

- [0123] 동작 850에서, 전자 장치(20)는 제1 액션을 수행한 후 마이크(예: 제2 마이크(222))를 통해 제2 사용자 발화를 수신할 수 있다. 전자 장치(20)는 웨이크업 발화를 수신한 후 선택된 시간 구간 동안 마이크를 통해 제2 사용자 발화를 수신할 수 있다.
- [0124] 동작 860에서, 전자 장치(20)는 ASR 모듈(예: 도 7의 ASR 모듈(720))을 이용하여 제2 사용자 발화에 대한 텍스트를 추출할 수 있다. 예를 들어, 전자 장치(20)는 제2 사용자 발화를 제2 음성 데이터로 변환하고 제2 음성 데이터를 텍스트로 변환할 수 있다. 전자 장치(20)는 제2 사용자 발화의 수신 시간 간격에 기반하여 종점을 검출하고, 종점에 기반하여 문장 단위의 텍스트를 추출할 수 있다.
- [0125] 동작 870에서, 전자 장치(20)는 제2 사용자 발화에 선택(지정)된 단어 또는 구문이 포함되어 있는지를 확인할 수 있다. 예를 들어, 전자 장치(20)는 각 문장에 포함된 텍스트와 선택된 단어 또는 구문을 비교하고, 비교 결과에 기반하여 각 문장에 선택된 단어 또는 구문이 포함되어 있는지를 확인할 수 있다.
- [0126] 동작 880에서, 제2 사용자 발화에 선택된 단어 또는 구문이 포함되어 있으면, 지능형 서버(10)는 제2 사용자 발화를 처리하여 제2 응답을 생성할 수 있다. 예를 들어, 제2 사용자 발화에 선택된 단어 또는 구문이 포함되어 있으면, 전자 장치(20)는 선택된 단어 또는 구문을 포함하는 문장을 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 NLU 모듈(750)을 이용하여, 선택된 단어 또는 구문을 포함하는 문장에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제2 액션을 결정할 수 있다. 지능형 서버(10)는 제2 액션 수행과 관련된 정보(제2 응답)를 생성하고, 생성된 제2 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 제2 액션 수행과 관련된 정보를 기반으로 제2 사용자 발화에 대응하는 제2 액션을 수행할 수 있다.
- [0127] 동작 870에서, 제2 사용자 발화에 선택된 단어 또는 구문이 포함되어 있지 않으면, 지능형 서버(10)는 제2 사용자 발화를 처리하여 제2 응답을 생성하지 않을 수 있다. 예를 들어, 전자 장치(20)는 제2 사용자 발화에 선택된 단어 또는 구문이 포함되어 있지 않으면, 제2 사용자 발화를 처리한 후 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다.
- [0128] 다양한 실시예에 따르면, 전자 장치(20)는 동작 810에서, 웨이크업 발화 대신에 입력 회로(210)를 통해 음성 기반 인텔리전트 어시스턴스 서비스를 호출하는 사용자 입력을 수신할 수 있다.
- [0130] 도 9는 일 실시 예에 따른 전자 장치에 의한 음성 인식 서비스 제공 방법(900)을 나타낸다.
- [0131] 도 9를 참조하면, 동작 910에서, 전자 장치(20)는 마이크(220)를 통해 웨이크업 발화를 수신하거나, 입력 회로(210)를 통해 지정된 입력을 수신할 수 있다.
- [0132] 동작 920에서, 전자 장치(20)는 웨이크업 발화 또는 지정된 입력을 수신하면, 음성 인식 서비스를 제공하는 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 실행할 수 있다.
- [0133] 동작 930에서, 전자 장치(20)는 마이크(220)를 통해 제1 사용자 발화를 수신할 수 있다. 전자 장치(20)는 제1 사용자 발화를 수신하면, 제1 사용자 발화를 제1 음성 데이터로 변환하여 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 NLU 모듈(750)을 이용하여, 제1 음성 데이터에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제1 액션을 결정할 수 있다. 지능형 서버(10)는 결정된 제1 액션 수행과 관련된 정보(제1 응답)를 생성하고, 생성된 제1 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다.
- [0134] 동작 940에서, 전자 장치(20)는 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행할 수 있다. 예를 들어, 전자 장치(20)는 지능형 서버(10)로부터 제1 액션 수행과 관련된 정보를 수신하고, 제1 액션 수행과 관련된 정보에 기반하여 제1 사용자 발화에 기반하여 결정된 제1 액션을 수행할 수 있다.
- [0135] 동작 950에서, 전자 장치(20)는 제1 액션을 수행한 시점으로부터 선택된 시간 내에 마이크(220)를 통해 제2 사용자 발화를 수신할 수 있다. 동작 950에서, 전자 장치(20)는 제2 사용자 발화를 수신하면서, 제2 사용자 발화를 실시간으로 제2 음성 데이터로 변환하고 제2 음성 데이터를 텍스트로 변환할 수 있다. 전자 장치(20)는 제2 사용자 발화의 수신 시간 간격에 기반하여 종점을 검출하고, 종점에 기반하여 문장 단위의 텍스트를 추출할 수

있다. 상기 선택된 시간은 상기 제1 응답이 생성된 이후부터의 시간을 포함할 수 있다.

- [0136] 동작 960에서, 전자 장치(20)는 제2 사용자 발화에 선택된 단어 또는 구문이 포함되어 있는지를 확인할 수 있다. 예를 들어, 전자 장치(20)는 제2 사용자 발화에 대한 텍스트(문장 단위의 텍스트)에 선택된 단어 또는 구문이 포함되는지를 확인할 수 있다.
- [0137] 동작 970에서, 전자 장치(20)는 액션을 수행(수행완료)한 시점으로부터 선택된 시간 내에 제2 사용자 발화에 기반하여 선택된 단어 또는 구문이 인식되면, 제2 사용자 발화에 기반하여 결정된 제2 액션을 수행할 수 있다. 예를 들어, 전자 장치(20)는 제2 사용자 발화에 대한 텍스트에 선택된 단어 또는 구문이 포함되어 있으면, 선택된 단어 또는 구문을 포함하는 문장을 지능형 서버(10)로 송신할 수 있다. 지능형 서버(10)는 NLU 모듈(750)을 이용하여, 선택된 단어 또는 구문을 포함하는 문장에 기반하여 사용자의 의도를 결정하고, 결정된 사용자의 의도에 따른 제2 액션을 결정할 수 있다. 지능형 서버(10)는 결정된 제2 액션 수행과 관련된 정보(제2 응답)를 생성하고, 생성된 제2 액션 수행과 관련된 정보를 전자 장치(20)로 송신할 수 있다. 전자 장치(20)는 제2 액션 수행과 관련된 정보를 수신하면, 제2 액션 수행과 관련된 정보에 기반하여 제2 사용자 발화에 대응하는 제2 액션을 수행할 수 있다.
- [0138] 동작 960에서, 제1 액션을 수행한 시점으로부터 선택된 시간 내에 제2 사용자 발화에 기반하여 선택된 단어 또는 구문이 인식되지 않으면, 전자 장치(20)는 동작 970에서 지능형 앱 또는 인텔리전트 어시스턴스 서비스를 종료할 수 있다.
- [0139] 상술한 실시 예에 따르면, 전자 장치(20)는 웨이크업 발화(또는, 지정된 입력)를 수신하고 음성 인식 서비스를 시작하면, 사용자 발화들에 트리거 단어가 포함되어 있는지를 실시간으로 모니터링하고, 웨이크업 발화를 대체하는 사용자 발화들의 일부(트리거 단어)에 기반하여 사용자 발화에 대응하는 액션을 수행할 수 있다. 따라서, 전자 장치(20)는 사용자가 복수의 액션들을 요청하기 위하여 웨이크업 발화를 반복적으로 수행하여야 하던 종래의 전자 장치의 불편을 해소할 수 있다.
- [0140] 도 10은, 다양한 실시예들에 따른, 네트워크 환경(1000) 내의 전자 장치(1001)의 블록도이다. 도 10을 참조하면, 네트워크 환경(1000)에서 전자 장치(1001)는 제 1 네트워크(1098)(예: 근거리 무선 통신 네트워크)를 통하여 전자 장치(1002)와 통신하거나, 또는 제 2 네트워크(1099)(예: 원거리 무선 통신 네트워크)를 통하여 전자 장치(1004) 또는 서버(1008)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(1001)는 서버(1008)를 통하여 전자 장치(1004)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(1001)는 프로세서(1020), 메모리(1030), 입력 장치(1050), 음향 출력 장치(1055), 표시 장치(1060), 오디오 모듈(1070), 센서 모듈(1076), 인터페이스(1077), 햅틱 모듈(1079), 카메라 모듈(1080), 전력 관리 모듈(1088), 배터리(1089), 통신 모듈(1090), 가입자 식별 모듈(1096), 또는 안테나 모듈(1097)을 포함할 수 있다. 어떤 실시예에서는, 전자 장치(1001)에는, 이 구성요소들 중 적어도 하나(예: 표시 장치(1060) 또는 카메라 모듈(1080))가 생략되거나, 하나 이상의 다른 구성요소가 추가될 수 있다. 어떤 실시예에서는, 이 구성요소들 중 일부들은 하나의 통합된 회로로 구현될 수 있다. 예를 들면, 센서 모듈(1076)(예: 지문 센서, 홍채 센서, 또는 조도 센서)은 표시 장치(1060)(예: 디스플레이)에 임베디드된 채 구현될 수 있다.
- [0141] 프로세서(1020)는, 예를 들면, 소프트웨어(예: 프로그램(1040))를 실행하여 프로세서(1020)에 연결된 전자 장치(1001)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 다양한 데이터 처리 또는 연산을 수행할 수 있다. 일실시예에 따르면, 데이터 처리 또는 연산의 적어도 일부로서, 프로세서(1020)는 다른 구성요소(예: 센서 모듈(1076) 또는 통신 모듈(1090))로부터 수신된 명령 또는 데이터를 휘발성 메모리(1032)에 로드하고, 휘발성 메모리(1032)에 저장된 명령 또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리(1034)에 저장할 수 있다. 일실시예에 따르면, 프로세서(1020)는 메인 프로세서(1021)(예: 중앙 처리 장치 또는 어플리케이션 프로세서), 및 이와는 독립적으로 또는 함께 운영 가능한 보조 프로세서(1023)(예: 그래픽 처리 장치, 이미지 시그널 프로세서, 센서 허브 프로세서, 또는 커뮤니케이션 프로세서)를 포함할 수 있다. 추가적으로 또는 대체적으로, 보조 프로세서(1023)은 메인 프로세서(1021)보다 저전력을 사용하거나, 또는 지정된 기능에 특화되도록 설정될 수 있다. 보조 프로세서(1023)는 메인 프로세서(1021)와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0142] 보조 프로세서(1023)는, 예를 들면, 메인 프로세서(1021)가 인액티브(예: 슬립) 상태에 있는 동안 메인 프로세서(1021)를 대신하여, 또는 메인 프로세서(1021)가 액티브(예: 어플리케이션 실행) 상태에 있는 동안 메인 프로세서(1021)와 함께, 전자 장치(1001)의 구성요소들 중 적어도 하나의 구성요소(예: 표시 장치(1060), 센서 모듈(1076), 또는 통신 모듈(1090))와 관련된 기능 또는 상태들의 적어도 일부를 제어할 수 있다. 일실시예에 따라

면, 보조 프로세서(1023)(예: 이미지 시그널 프로세서 또는 커뮤니케이션 프로세서)는 기능적으로 관련 있는 다른 구성 요소(예: 카메라 모듈(1080) 또는 통신 모듈(1090))의 일부로서 구현될 수 있다.

- [0143] 메모리(1030)는, 전자 장치(1001)의 적어도 하나의 구성요소(예: 프로세서(1020) 또는 센서모듈(1076))에 의해 사용되는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(예: 프로그램(1040)) 및, 이와 관련된 명령에 대한 입력 데이터 또는 출력 데이터를 포함할 수 있다. 메모리(1030)는, 휘발성 메모리(1032) 또는 비휘발성 메모리(1034)를 포함할 수 있다.
- [0144] 프로그램(1040)은 메모리(1030)에 소프트웨어로서 저장될 수 있으며, 예를 들면, 운영 체제(1042), 미들 웨어(1044) 또는 어플리케이션(1046)을 포함할 수 있다.
- [0145] 입력 장치(1050)는, 전자 장치(1001)의 구성요소(예: 프로세서(1020))에 사용될 명령 또는 데이터를 전자 장치(1001)의 외부(예: 사용자)로부터 수신할 수 있다. 입력 장치(1050)은, 예를 들면, 마이크, 마우스, 키보드, 또는 디지털 펜(예:스타일러스 펜)을 포함할 수 있다.
- [0146] 음향 출력 장치(1055)는 음향 신호를 전자 장치(1001)의 외부로 출력할 수 있다. 음향 출력 장치(1055)는, 예를 들면, 스피커 또는 리시버를 포함할 수 있다. 스피커는 멀티미디어 재생 또는 녹음 재생과 같이 일반적인 용도로 사용될 수 있고, 리시버는 착신 전화를 수신하기 위해 사용될 수 있다. 일실시예에 따르면, 리시버는 스피커와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0147] 표시 장치(1060)는 전자 장치(1001)의 외부(예: 사용자)로 정보를 시각적으로 제공할 수 있다. 표시 장치(1060)은, 예를 들면, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 일실시예에 따르면, 표시 장치(1060)는 터치를 감지하도록 설정된 터치 회로(touch circuitry), 또는 상기 터치에 의해 발생하는 힘의 세기를 측정하도록 설정된 센서 회로(예: 압력 센서)를 포함할 수 있다.
- [0148] 오디오 모듈(1070)은 소리를 전기 신호로 변환시키거나, 반대로 전기 신호를 소리로 변환시킬 수 있다. 일실시예에 따르면, 오디오 모듈(1070)은, 입력 장치(1050)를 통해 소리를 획득하거나, 음향 출력 장치(1055), 또는 전자 장치(1001)와 직접 또는 무선으로 연결된 외부 전자 장치(예: 전자 장치(1002)) (예: 스피커 또는 헤드폰))를 통해 소리를 출력할 수 있다.
- [0149] 센서 모듈(1076)은 전자 장치(1001)의 작동 상태(예: 전력 또는 온도), 또는 외부의 환경 상태(예: 사용자 상태)를 감지하고, 감지된 상태에 대응하는 전기 신호 또는 데이터 값을 생성할 수 있다. 일실시예에 따르면, 센서 모듈(1076)은, 예를 들면, 제스처 센서, 자이로 센서, 기압 센서, 마그네틱 센서, 가속도 센서, 그립 센서, 근접 센서, 컬러 센서, IR(infrared) 센서, 생체 센서, 온도 센서, 습도 센서, 또는 조도 센서를 포함할 수 있다.
- [0150] 인터페이스(1077)는 전자 장치(1001)이 외부 전자 장치(예: 전자 장치(1002))와 직접 또는 무선으로 연결되기 위해 사용될 수 있는 하나 이상의 지정된 프로토콜들을 지원할 수 있다. 일실시예에 따르면, 인터페이스(1077)는, 예를 들면, HDMI(high definition multimedia interface), USB(universal serial bus) 인터페이스, SD카드 인터페이스, 또는 오디오 인터페이스를 포함할 수 있다.
- [0151] 연결 단자(1078)는, 그를 통해서 전자 장치(1001)가 외부 전자 장치(예: 전자 장치(1002))와 물리적으로 연결될 수 있는 커넥터를 포함할 수 있다. 일실시예에 따르면, 연결 단자(1078)은, 예를 들면, HDMI 커넥터, USB 커넥터, SD 카드 커넥터, 또는 오디오 커넥터(예: 헤드폰 커넥터)를 포함할 수 있다.
- [0152] 햅틱 모듈(1079)은 전기적 신호를 사용자가 촉각 또는 운동 감각을 통해서 인지할 수 있는 기계적인 자극(예: 진동 또는 움직임) 또는 전기적인 자극으로 변환할 수 있다. 일실시예에 따르면, 햅틱 모듈(1079)은, 예를 들면, 모터, 압전 소자, 또는 전기 자극 장치를 포함할 수 있다.
- [0153] 카메라 모듈(1080)은 정지 영상 및 동영상을 촬영할 수 있다. 일실시예에 따르면, 카메라 모듈(1080)은 하나 이상의 렌즈들, 이미지 센서들, 이미지 시그널 프로세서들, 또는 플래시들을 포함할 수 있다.
- [0154] 전력 관리 모듈(1088)은 전자 장치(1001)에 공급되는 전력을 관리할 수 있다. 일실시예에 따르면, 전력 관리 모듈(1088)은, 예를 들면, PMIC(power management integrated circuit)의 적어도 일부로서 구현될 수 있다.
- [0155] 배터리(1089)는 전자 장치(1001)의 적어도 하나의 구성 요소에 전력을 공급할 수 있다. 일실시예에 따르면, 배터리(1089)는, 예를 들면, 재충전 불가능한 1차 전지, 재충전 가능한 2차 전지 또는 연료 전지를 포함할 수 있다.

- [0156] 통신 모듈(1090)은 전자 장치(1001)와 외부 전자 장치(예: 전자 장치(1002), 전자 장치(1004), 또는 서버(1008))간의 직접(예: 유선) 통신 채널 또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신 모듈(1090)은 프로세서(1020)(예: 어플리케이션 프로세서)와 독립적으로 운영되고, 직접(예: 유선) 통신 또는 무선 통신을 지원하는 하나 이상의 커뮤니케이션 프로세서를 포함할 수 있다. 일실시예에 따르면, 통신 모듈(1090)은 무선 통신 모듈(1092)(예: 셀룰러 통신 모듈, 근거리 무선 통신 모듈, 또는 GNSS(global navigation satellite system) 통신 모듈) 또는 유선 통신 모듈(1094)(예: LAN(local area network) 통신 모듈, 또는 전력선 통신 모듈)을 포함할 수 있다. 이들 통신 모듈 중 해당하는 통신 모듈은 제 1 네트워크(1098)(예: 블루투스, WiFi direct 또는 IrDA(infrared data association) 같은 근거리 통신 네트워크) 또는 제 2 네트워크(1099)(예: 셀룰러 네트워크, 인터넷, 또는 컴퓨터 네트워크(예: LAN 또는 WAN)와 같은 원거리 통신 네트워크)를 통하여 외부 전자 장치와 통신할 수 있다. 이런 여러 종류의 통신 모듈들은 하나의 구성 요소(예: 단일 칩)으로 통합되거나, 또는 서로 별도의 복수의 구성 요소들(예: 복수 칩들)로 구현될 수 있다. 무선 통신 모듈(1092)은 가입자 식별 모듈(1096)에 저장된 가입자 정보(예: 국제 모바일 가입자 식별자(IMSI))를 이용하여 제 1 네트워크(1098) 또는 제 2 네트워크(1099)와 같은 통신 네트워크 내에서 전자 장치(1001)를 확인 및 인증할 수 있다.
- [0157] 안테나 모듈(1097)은 신호 또는 전력을 외부(예: 외부 전자 장치)로 송신하거나 외부로부터 수신할 수 있다. 일실시예에 따르면, 안테나 모듈은 서브스트레이트(예: PCB) 위에 형성된 도전체 또는 도전성 패턴으로 이루어진 방사체를 포함하는 하나의 안테나를 포함할 수 있다. 일실시예에 따르면, 안테나 모듈(1097)은 복수의 안테나들을 포함할 수 있다. 이런 경우, 제 1 네트워크(1098) 또는 제 2 네트워크(1099)와 같은 통신 네트워크에서 사용되는 통신 방식에 적합한 적어도 하나의 안테나가, 예를 들면, 통신 모듈(1090)에 의하여 상기 복수의 안테나들로부터 선택될 수 있다. 신호 또는 전력은 상기 선택된 적어도 하나의 안테나를 통하여 통신 모듈(1090)과 외부 전자 장치 간에 송신되거나 수신될 수 있다. 어떤 실시예에 따르면, 방사체 이외에 다른 부품(예: RFIC)이 추가로 안테나 모듈(1097)의 일부로 형성될 수 있다.
- [0158] 상기 구성요소들 중 적어도 일부는 주변 기기들간 통신 방식(예: 버스, GPIO(general purpose input and output), SPI(serial peripheral interface), 또는 MIPI(mobile industry processor interface))를 통해 서로 연결되고 신호(예: 명령 또는 데이터)를 상호간에 교환할 수 있다.
- [0159] 일실시예에 따르면, 명령 또는 데이터는 제 2 네트워크(1099)에 연결된 서버(1008)를 통해서 전자 장치(1001)와 외부의 전자 장치(1004)간에 송신 또는 수신될 수 있다. 전자 장치(1002, 1004) 각각은 전자 장치(1001)와 동일한 또는 다른 종류의 장치일 수 있다. 일실시예에 따르면, 전자 장치(1001)에서 실행되는 동작들의 전부 또는 일부는 외부 전자 장치들(1002, 1004, or 1008) 중 하나 이상의 외부 장치들에서 실행될 수 있다. 예를 들면, 전자 장치(1001)가 어떤 기능이나 서비스를 자동으로, 또는 사용자 또는 다른 장치로부터의 요청에 반응하여 수행해야 할 경우에, 전자 장치(1001)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 하나 이상의 외부 전자 장치들에게 그 기능 또는 그 서비스의 적어도 일부를 수행하라고 요청할 수 있다. 상기 요청을 수신한 하나 이상의 외부 전자 장치들은 요청된 기능 또는 서비스의 적어도 일부, 또는 상기 요청과 관련된 추가 기능 또는 서비스를 실행하고, 그 실행의 결과를 전자 장치(1001)로 전달할 수 있다. 전자 장치(1001)는 상기 결과를, 그대로 또는 추가적으로 처리하여, 상기 요청에 대한 응답의 적어도 일부로서 제공할 수 있다. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.
- [0160] 도 11은 일 실시예에 따른 통합 지능 (integrated intelligence) 시스템을 나타낸 블록도이다.
- [0161] 도 11을 참조하면, 일 실시예의 통합 지능 시스템(110)(예: 도 1의 통합 지능 시스템(10, 20)은 사용자 단말(1100)(예: 도 2의 전자 장치(20)), 지능형 서버(1200)(예: 도 1의 지능형 서버(10)), 및 서비스 서버(1300)를 포함할 수 있다.
- [0162] 일 실시 예의 사용자 단말(1100)은, 인터넷에 연결 가능한 단말 장치(또는, 전자 장치)일 수 있으며, 예를 들어, 휴대폰, 스마트폰, PDA(personal digital assistant), 노트북 컴퓨터, TV, 백색 가전, 웨어러블 장치, HMD, 또는 스마트 스피커일 수 있다.
- [0163] 도시된 실시 예에 따르면, 사용자 단말(1100)은 통신 인터페이스(1110)(예: 도 2의 통신 회로(230)), 마이크(1120)(예: 도 2의 마이크(220)), 스피커(1130)(예: 도 2의 스피커(240)), 디스플레이(1140)(예: 도 2의 디스플레이(210)), 메모리(1150)(예: 도 2의 메모리(250)), 또는 프로세서(1160)(예: 도 2의 프로세서(260))를 포함할

수 있다. 상기 열거된 구성요소들은 서로 작동적으로 또는 전기적으로 연결될 수 있다.

- [0164] 일 실시 예의 통신 인터페이스(1110)는 외부 장치와 연결되어 데이터를 송수신하도록 구성될 수 있다. 일 실시 예의 마이크(1120)는 소리(예: 사용자 발화)를 수신하여, 전기적 신호로 변환할 수 있다. 일 실시 예의 스피커(1130)는 전기적 신호를 소리(예: 음성)으로 출력할 수 있다. 일 실시 예의 디스플레이(1140)는 이미지 또는 비디오를 표시하도록 구성될 수 있다. 일 실시 예의 디스플레이(1140)는 또한 실행되는 앱(app)(또는, 어플리케이션 프로그램(application program))의 그래픽 사용자 인터페이스(graphic user interface)(GUI)를 표시할 수 있다.
- [0165] 일 실시 예의 메모리(1150)는 클라이언트 모듈(1151), SDK(software development kit)(1153), 및 복수의 앱들(1155)을 저장할 수 있다. 상기 클라이언트 모듈(1151), 및 SDK(1153)는 범용적인 기능을 수행하기 위한 프레임워크(framework)(또는, 솔루션 프로그램)를 구성할 수 있다. 또한, 클라이언트 모듈(1151) 또는 SDK(1153)는 음성 입력을 처리하기 위한 프레임워크를 구성할 수 있다.
- [0166] 일 실시 예의 메모리(1150)는 상기 복수의 앱들(1155)은 지정된 기능을 수행하기 위한 프로그램일 수 있다. 일 실시 예에 따르면, 복수의 앱(1155)은 제1 앱(1155_1), 제2 앱(1155_3)을 포함할 수 있다. 일 실시 예에 따르면, 복수의 앱(1155) 각각은 지정된 기능을 수행하기 위한 복수의 동작들을 포함할 수 있다. 예를 들어, 상기 앱들은, 알람 앱, 메시지 앱, 및/또는 스케줄 앱을 포함할 수 있다. 일 실시 예에 따르면, 복수의 앱들(1155)은 프로세서(1160)에 의해 실행되어 상기 복수의 동작들 중 적어도 일부를 순차적으로 실행할 수 있다.
- [0167] 일 실시 예의 프로세서(1160)는 사용자 단말(1100)의 전반적인 동작을 제어할 수 있다. 예를 들어, 프로세서(1160)는 통신 인터페이스(1110), 마이크(1120), 스피커(1130), 및 디스플레이(1140)와 전기적으로 연결되어 연결되어 지정된 동작을 수행할 수 있다.
- [0168] 일 실시 예의 프로세서(1160)는 또한 상기 메모리(1150)에 저장된 프로그램을 실행시켜 지정된 기능을 수행할 수 있다. 예를 들어, 프로세서(1160)는 클라이언트 모듈(1151) 또는 SDK(1153) 중 적어도 하나를 실행하여, 음성 입력을 처리하기 위한 이하의 동작을 수행할 수 있다. 프로세서(1160)는, 예를 들어, SDK(1153)를 통해 복수의 앱(1155)의 동작을 제어할 수 있다. 클라이언트 모듈(1151) 또는 SDK(1153)의 동작으로 설명된 이하의 동작은 프로세서(1160)의 실행에 의한 동작일 수 있다.
- [0169] 일 실시 예의 클라이언트 모듈(1151)은 음성 입력을 수신할 수 있다. 예를 들어, 클라이언트 모듈(1151)은 마이크(1120)를 통해 감지된 사용자 발화에 대응되는 음성 신호를 수신할 수 있다. 상기 클라이언트 모듈(1151)은 수신된 음성 입력을 지능형 서버(1200)로 송신할 수 있다. 클라이언트 모듈(1151)은 수신된 음성 입력과 함께, 사용자 단말(1100)의 상태 정보를 지능형 서버(1200)로 송신할 수 있다. 상기 상태 정보는, 예를 들어, 앱의 실행 상태 정보일 수 있다.
- [0170] 일 실시 예의 클라이언트 모듈(1151)은 수신된 음성 입력에 대응되는 결과를 수신할 수 있다. 예를 들어, 클라이언트 모듈(1151)은 지능형 서버(1200)에서 상기 수신된 음성 입력에 대응되는 결과를 산출할 수 있는 경우, 수신된 음성 입력에 대응되는 결과를 수신할 수 있다. 클라이언트 모듈(1151)은 상기 수신된 결과를 디스플레이(1140)에 표시할 수 있다.
- [0171] 일 실시 예의 클라이언트 모듈(1151)은 수신된 음성 입력에 대응되는 플랜을 수신할 수 있다. 클라이언트 모듈(1151)은 플랜에 따라 앱의 복수의 동작을 실행한 결과를 디스플레이(1140)에 표시할 수 있다. 클라이언트 모듈(1151)은, 예를 들어, 복수의 동작의 실행 결과를 순차적으로 디스플레이에 표시할 수 있다. 사용자 단말(1100)은, 다른 예를 들어, 복수의 동작을 실행한 일부 결과(예: 마지막 동작의 결과)만을 디스플레이에 표시할 수 있다.
- [0172] 일 실시 예에 따르면, 클라이언트 모듈(1151)은 지능형 서버(1200)로부터 음성 입력에 대응되는 결과를 산출하기 위해 필요한 정보를 획득하기 위한 요청을 수신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(1151)은 상기 요청에 대응하여 상기 필요한 정보를 지능형 서버(1200)로 송신할 수 있다.
- [0173] 일 실시 예의 클라이언트 모듈(1151)은 플랜에 따라 복수의 동작을 실행한 결과 정보를 지능형 서버(1200)로 송신할 수 있다. 지능형 서버(1200)는 상기 결과 정보를 이용하여 수신된 음성 입력이 올바르게 처리된 것을 확인할 수 있다.
- [0174] 일 실시 예의 클라이언트 모듈(1151)은 음성 인식 모듈을 포함할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(1151)은 상기 음성 인식 모듈을 통해 제한된 기능을 수행하는 음성 입력을 인식할 수 있다. 예를 들어, 클라

이언트 모듈(1151)은 지정된 입력(예: 웨이크 업!)을 통해 유기적인 동작을 수행하기 위한 음성 입력을 처리하기 위한 지능형 앱(지능형 앱 또는 인텔리전트 어시스턴스 서비스)을 수행할 수 있다.

- [0175] 일 실시 예의 지능형 서버(1200)는 통신 망을 통해 사용자 단말(1100)로부터 사용자 음성 입력과 관련된 정보를 수신할 수 있다. 일 실시 예에 따르면, 지능형 서버(1200)는 수신된 음성 입력과 관련된 데이터를 텍스트 데이터(text data)로 변경할 수 있다. 일 실시 예에 따르면, 지능형 서버(1200)는 상기 텍스트 데이터에 기초하여 사용자 음성 입력과 대응되는 태스크(task)를 수행하기 위한 플랜(plan)을 생성할 수 있다
- [0176] 일 실시 예에 따르면, 플랜은 인공 지능(artificial intelligent)(AI) 시스템에 의해 생성될 수 있다. 인공지능 시스템은 룰 베이스 시스템(rule-based system) 일 수도 있고, 신경망 베이스 시스템(neural network-based system)(예: 피드포워드 신경망(feedforward neural network(FNN)), 순환 신경망(recurrent neural network(RNN))) 일 수도 있다. 또는, 전술한 것의 조합 또는 이와 다른 인공지능 시스템일 수도 있다. 일 실시 예에 따르면, 플랜은 미리 정의된 플랜의 집합에서 선택될 수 있거나, 사용자 요청에 응답하여 실시간으로 생성될 수 있다. 예를 들어, 인공지능 시스템은 미리 정의된 복수의 플랜 중 적어도 플랜을 선택할 수 있다.
- [0177] 일 실시 예의 지능형 서버(1200)는 생성된 플랜에 따른 결과를 사용자 단말(1100)로 송신하거나, 생성된 플랜을 사용자 단말(1100)로 송신할 수 있다. 일 실시 예에 따르면, 사용자 단말(1100)은 플랜에 따른 결과를 디스플레이에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(1100)은 플랜에 따른 동작을 실행한 결과를 디스플레이에 표시할 수 있다.
- [0178] 일 실시 예의 지능형 서버(1200)는 프론트 엔드(front end)(1210), 자연어 플랫폼(natural language platform)(1220), 캡슐 데이터베이스(capsule DB)(1230), 실행 엔진(execution engine)(1240), 엔드 사용자 인터페이스(end user interface)(1250), 매니지먼트 플랫폼(management platform)(1260), 빅 데이터 플랫폼(big data platform)(1270), 또는 분석 플랫폼(analytic platform)(1280)을 포함할 수 있다.
- [0179] 일 실시 예의 프론트 엔드(1210)는 사용자 단말(1100)로부터 수신된 음성 입력을 수신할 수 있다. 프론트 엔드(1210)는 상기 음성 입력에 대응되는 응답을 송신할 수 있다.
- [0180] 일 실시 예에 따르면, 자연어 플랫폼(1220)은 자동 음성 인식 모듈(automatic speech recognition module)(ASR module)(1221), 자연어 이해 모듈(natural language understanding module)(NLU module)(1223), 플래너 모듈(planner module)(1225), 자연어 생성 모듈(natural language generator module)(NLG module)(1227)또는 텍스트 음성 변환 모듈(text to speech module)(TTS module)(1229)를 포함할 수 있다.
- [0181] 일 실시 예의 자동 음성 인식 모듈(1221)은 사용자 단말(1100)로부터 수신된 음성 입력을 텍스트 데이터로 변환할 수 있다. 일 실시 예의 자연어 이해 모듈(1223)은 음성 입력의 텍스트 데이터를 이용하여 사용자의 의도를 파악할 수 있다. 예를 들어, 자연어 이해 모듈(1223)은 문법적 분석(syntactic analyze) 또는 의미적 분석(semantic analyze)을 수행하여 사용자의 의도를 파악할 수 있다. 일 실시 예의 자연어 이해 모듈(1223)은 형태소 또는 구의 언어적 특징(예: 문법적 요소)을 이용하여 음성 입력으로부터 추출된 단어의 의미를 파악하고, 상기 파악된 단어의 의미를 의도에 매칭시켜 사용자의 의도를 결정할 수 있다.
- [0182] 일 실시 예의 플래너 모듈(1225)은 자연어 이해 모듈(1223)에서 결정된 의도 및 파라미터를 이용하여 플랜을 생성할 수 있다. 일 실시 예에 따르면, 플래너 모듈(1225)은 상기 결정된 의도에 기초하여 태스크를 수행하기 위해 필요한 복수의 도메인을 결정할 수 있다. 플래너 모듈(1225)은 상기 의도에 기초하여 결정된 복수의 도메인 각각에 포함된 복수의 동작을 결정할 수 있다. 일 실시 예에 따르면, 플래너 모듈(1225)은 상기 결정된 복수의 동작을 실행하는데 필요한 파라미터나, 상기 복수의 동작의 실행에 의해 출력되는 결과 값을 결정할 수 있다. 상기 파라미터, 및 상기 결과 값은 지정된 형식(또는, 클래스)의 컨셉으로 정의될 수 있다. 이에 따라, 플랜은 사용자의 의도에 의해 결정된 복수의 동작, 및 복수의 컨셉을 포함할 수 있다. 상기 플래너 모듈(1225)은 상기 복수의 동작, 및 상기 복수의 컨셉 사이의 관계를 단계적(또는, 계층적)으로 결정할 수 있다. 예를 들어, 플래너 모듈(1225)은 복수의 컨셉에 기초하여 사용자의 의도에 기초하여 결정된 복수의 동작의 실행 순서를 결정할 수 있다. 다시 말해, 플래너 모듈(1225)은 복수의 동작의 실행에 필요한 파라미터, 및 복수의 동작의 실행에 의해 출력되는 결과에 기초하여, 복수의 동작의 실행 순서를 결정할 수 있다. 이에 따라, 플래너 모듈(1225)은 복수의 동작, 및 복수의 컨셉 사이의 연관 정보(예: 온톨로지(ontology))가 포함된 플랜을 생성할 수 있다. 상기 플래너 모듈(1225)은 컨셉과 동작의 관계들의 집합이 저장된 캡슐 데이터베이스(1230)에 저장된 정보를 이용하여 플랜을 생성할 수 있다.
- [0183] 일 실시 예의 자연어 생성 모듈(1227)은 지정된 정보를 텍스트 형태로 변경할 수 있다. 상기 텍스트 형태로 변

경된 정보는 자연어 발화의 형태일 수 있다. 일 실시 예의 텍스트 음성 변환 모듈(1229)은 텍스트 형태의 정보를 음성 형태의 정보로 변경할 수 있다.

[0184] 일 실시 예에 따르면, 자연어 플랫폼(1220)의 기능의 일부 기능 또는 전체 기능은 사용자 단말(1100)에서도 구현가능 할 수 있다.

[0185] 상기 캡슐 데이터베이스(1230)는 복수의 도메인에 대응되는 복수의 컨셉과 동작들의 관계에 대한 정보를 저장할 수 있다. 일 실시예에 따른 캡슐은 플랜에 포함된 복수의 동작 오브젝트(action object 또는, 동작 정보) 및 컨셉 오브젝트(concept object 또는 컨셉 정보)를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(1230)는 CAN(concept action network)의 형태로 복수의 캡슐을 저장할 수 있다. 일 실시 예에 따르면, 복수의 캡슐은 캡슐 데이터베이스(1230)에 포함된 기능 저장소(function registry)에 저장될 수 있다.

[0186] 상기 캡슐 데이터베이스(1230)는 음성 입력에 대응되는 플랜을 결정할 때 필요한 전략 정보가 저장된 전략 레지스트리(strategy registry)를 포함할 수 있다. 상기 전략 정보는 음성 입력에 대응되는 복수의 플랜이 있는 경우, 하나의 플랜을 결정하기 위한 기준 정보를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(1230)는 지정된 상황에서 사용자에게 후속 동작을 제안하기 위한 후속 동작의 정보가 저장된 후속 동작 레지스트리(follow up registry)를 포함할 수 있다. 상기 후속 동작은, 예를 들어, 후속 발화를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(1230)는 사용자 단말(1100)을 통해 출력되는 정보의 레이아웃(layout) 정보를 저장하는 레이아웃 레지스트리(layout registry)를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(1230)는 캡슐 정보에 포함된 어휘(vocabulary) 정보가 저장된 어휘 레지스트리(vocabulary registry)를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(1230)는 사용자와의 대화(dialog)(또는, 인터랙션(interaction)) 정보가 저장된 대화 레지스트리(dialog registry)를 포함할 수 있다. 상기 캡슐 데이터베이스(1230)는 개발자 툴(developer tool)을 통해 저장된 오브젝트를 업데이트(update)할 수 있다. 상기 개발자 툴은, 예를 들어, 동작 오브젝트 또는 컨셉 오브젝트를 업데이트하기 위한 기능 에디터(function editor)를 포함할 수 있다. 상기 개발자 툴은 어휘를 업데이트하기 위한 어휘 에디터(vocabulary editor)를 포함할 수 있다. 상기 개발자 툴은 플랜을 결정하는 전략을 생성 및 등록 하는 전략 에디터(strategy editor)를 포함할 수 있다. 상기 개발자 툴은 사용자와의 대화를 생성하는 대화 에디터(dialog editor)를 포함할 수 있다. 상기 개발자 툴은 후속 목표를 활성화하고, 힌트를 제공하는 후속 발화를 편집할 수 있는 후속 동작 에디터(follow up editor)를 포함할 수 있다. 상기 후속 목표는 현재 설정된 목표, 사용자의 선호도 또는 환경 조건에 기초하여 결정될 수 있다. 일 실시 예에서는 캡슐 데이터베이스(1230)은 사용자 단말(1100) 내에도 구현이 가능할 수 있다.

[0187] 일 실시 예의 실행 엔진(1240)은 상기 생성된 플랜을 이용하여 결과를 산출할 수 있다. 엔드 사용자 인터페이스(1250)는 산출된 결과를 사용자 단말(1100)로 송신할 수 있다. 이에 따라, 사용자 단말(1100)은 상기 결과를 수신하고, 상기 수신된 결과를 사용자에게 제공할 수 있다. 일 실시 예의 매니지먼트 플랫폼(1260)은 지능형 서버(1200)에서 이용되는 정보를 관리할 수 있다. 일 실시 예의 빅 데이터 플랫폼(1270)은 사용자의 데이터를 수집할 수 있다. 일 실시 예의 분석 플랫폼(1280)을 지능형 서버(1200)의 QoS(quality of service)를 관리할 수 있다. 예를 들어, 분석 플랫폼(1280)은 지능형 서버(1200)의 구성 요소 및 처리 속도(또는, 효율성)를 관리할 수 있다.

[0188] 일 실시 예의 서비스 서버(1300)는 사용자 단말(1100)에 지정된 서비스(예: 음식 주문 또는 호텔 예약)를 제공할 수 있다. 일 실시 예에 따르면, 서비스 서버(1300)는 제3 자에 의해 운영되는 서버일 수 있다. 일 실시 예의 서비스 서버(1300)는 수신된 음성 입력에 대응되는 플랜을 생성하기 위한 정보를 지능형 서버(1200)에 제공할 수 있다. 상기 제공된 정보는 캡슐 데이터베이스(1230)에 저장될 수 있다. 또한, 서비스 서버(1300)는 플랜에 따른 결과 정보를 지능형 서버(1200)에 제공할 수 있다.

[0189] 위에 기술된 통합 지능 시스템(110)에서, 상기 사용자 단말(1100)은, 사용자 입력에 응답하여 사용자에게 다양한 인텔리전트 서비스를 제공할 수 있다. 상기 사용자 입력은, 예를 들어, 물리적 버튼을 통한 입력, 터치 입력 또는 음성 입력을 포함할 수 있다.

[0190] 일 실시 예에서, 상기 사용자 단말(1100)은 내부에 저장된 지능형 앱(또는, 음성 인식 앱)을 통해 음성 인식 서비스를 제공할 수 있다. 이 경우, 예를 들어, 사용자 단말(1100)은 상기 마이크를 통해 수신된 사용자 발화(utterance) 또는 음성 입력(voice input)를 인식하고, 인식된 음성 입력에 대응되는 서비스를 사용자에게 제공할 수 있다.

[0191] 일 실시 예에서, 사용자 단말(1100)은 수신된 음성 입력에 기초하여, 단독으로 또는 상기 지능형 서버 및/또는

서비스 서버와 함께 지정된 동작을 수행할 수 있다. 예를 들어, 사용자 단말(1100)은 수신된 음성 입력에 대응되는 앱을 실행시키고, 실행된 앱을 통해 지정된 동작을 수행할 수 있다.

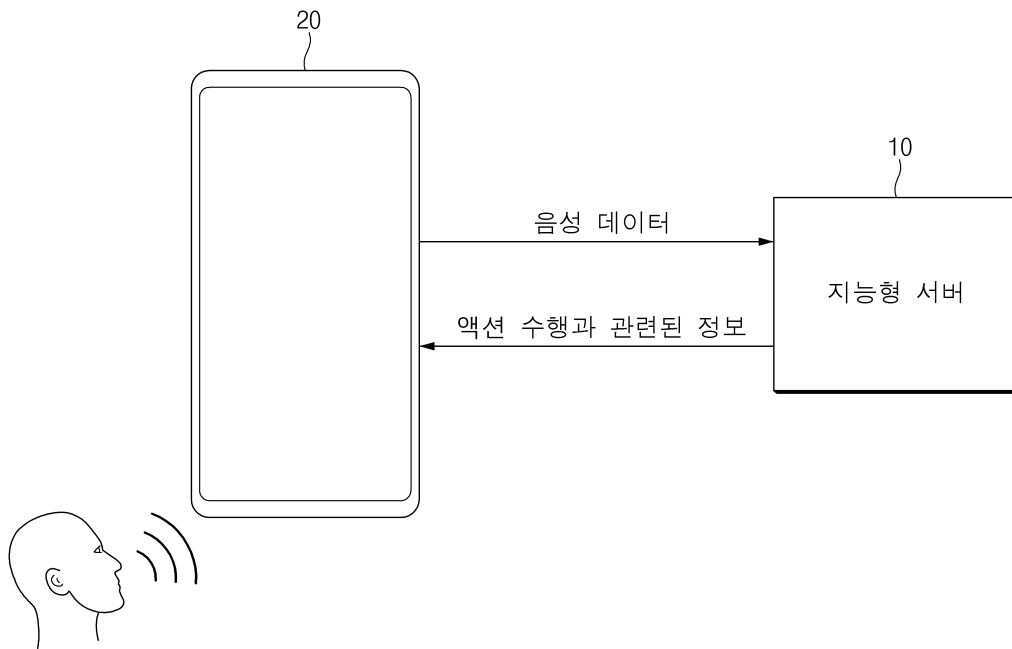
- [0192] 일 실시 예에서, 사용자 단말(1100)이 지능형 서버(1200) 및/또는 서비스 서버와 함께 서비스를 제공하는 경우에는, 상기 사용자 단말은, 상기 마이크(1120)를 이용하여 사용자 발화를 감지하고, 상기 감지된 사용자 발화에 대응되는 신호(또는, 음성 데이터)를 생성할 수 있다. 상기 사용자 단말은, 상기 음성 데이터를 통신 인터페이스(1110)를 이용하여 지능형 서버(1200)로 송신할 수 있다.
- [0193] 일 실시 예에 따른 지능형 서버(1200)는 사용자 단말(1100)로부터 수신된 음성 입력에 대한 응답으로써, 음성 입력에 대응되는 태스크(task)를 수행하기 위한 플랜, 또는 상기 플랜에 따라 동작을 수행한 결과를 생성할 수 있다. 상기 플랜은, 예를 들어, 사용자의 음성 입력에 대응되는 태스크(task)를 수행하기 위한 복수의 동작, 및 상기 복수의 동작과 관련된 복수의 컨셉을 포함할 수 있다. 상기 컨셉은 상기 복수의 동작의 실행에 입력되는 파라미터나, 복수의 동작의 실행에 의해 출력되는 결과 값을 정의한 것일 수 있다. 상기 플랜은 복수의 동작, 및 복수의 컨셉 사이의 연관 정보를 포함할 수 있다.
- [0194] 일 실시 예의 사용자 단말(1100)은, 통신 인터페이스(1110)를 이용하여 상기 응답을 수신할 수 있다. 사용자 단말(1100)은 상기 스피커(1130)를 이용하여 사용자 단말(1100) 내부에서 생성된 음성 신호를 외부로 출력하거나, 디스플레이(1140)를 이용하여 사용자 단말(1100) 내부에서 생성된 이미지를 외부로 출력할 수 있다.
- [0195] 도 12는 다양한 실시 예에 따른, 컨셉과 동작의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.
- [0196] 상기 지능형 서버(1200)의 캡슐 데이터베이스(예: 캡슐 데이터베이스(1230))는 CAN (concept action network) 형태로 캡슐을 저장할 수 있다. 상기 캡슐 데이터베이스는 사용자의 음성 입력에 대응되는 태스크를 처리하기 위한 동작, 및 상기 동작을 위해 필요한 파라미터를 CAN(concept action network) 형태로 저장될 수 있다.
- [0197] 상기 캡슐 데이터베이스는 복수의 도메인(예: 어플리케이션) 각각에 대응되는 복수의 캡슐(capsule(A)(4010), capsule(B)(4040))을 저장할 수 있다. 일 실시 예에 따르면, 하나의 캡슐(예: capsule(A)(4010))은 하나의 도메인(예: 위치(geo), 어플리케이션)에 대응될 수 있다. 또한, 하나의 캡슐에는 캡슐과 관련된 도메인에 대한 기능을 수행하기 위한 적어도 하나의 서비스 제공자(예: CP 1(4020) 또는 CP 2 (4030))가 대응될 수 있다. 일 실시 예에 따르면, 하나의 캡슐은 지정된 기능을 수행하기 위한 적어도 하나 이상의 동작(4100) 및 적어도 하나 이상의 컨셉(4200)을 포함할 수 있다.
- [0198] 상기, 자연어 플랫폼(1220)은 캡슐 데이터베이스에 저장된 캡슐을 이용하여 수신된 음성 입력에 대응하는 태스크를 수행하기 위한 플랜을 생성할 수 있다. 예를 들어, 자연어 플랫폼의 플래너 모듈(1225)은 캡슐 데이터베이스에 저장된 캡슐을 이용하여 플랜을 생성할 수 있다. 예를 들어, 캡슐 A (4010)의 동작들(4011,4013)과 컨셉들(4012,4014) 및 캡슐 B(4004)의 동작(4041)과 컨셉(4042)를 이용하여 플랜(4070)을 생성할 수 있다.
- [0199] 도 13는 다양한 실시 예에 따른 사용자 단말이 지능형 앱(지능형 앱 또는)을 통해 수신된 음성 입력을 처리하는 화면을 나타낸 도면이다.
- [0200] 사용자 단말(1100)은 지능형 서버(1200)를 통해 사용자 입력을 처리하기 위해 지능형 앱을 실행할 수 있다.
- [0201] 일 실시 예에 따르면, 3100 화면에서, 사용자 단말(1100)은 지정된 음성 입력(예: 웨이크 업!)를 인식하거나 하드웨어 키(예: 전용 하드웨어 키)를 통한 입력을 수신하면, 음성 입력을 처리하기 위한 지능형 앱을 실행할 수 있다. 사용자 단말(1100)은, 예를 들어, 스케줄 앱을 실행한 상태에서 지능형 앱을 실행할 수 있다. 일 실시 예에 따르면, 사용자 단말(1100)은 지능형 앱에 대응되는 오브젝트(예: 아이콘)(3110)를 디스플레이(1140)에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(1100)은 사용자 발화에 의한 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(1100)은 "이번주 일정 알려줘!"라는 음성 입력을 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말(1100)은 수신된 음성 입력의 텍스트 데이터가 표시된 지능형 앱의 UI(user interface)(3130)(예: 입력창)를 디스플레이에 표시할 수 있다.
- [0202] 일 실시 예에 따르면, 3200 화면에서, 사용자 단말(1100)은 수신된 음성 입력에 대응되는 결과를 디스플레이에 표시할 수 있다. 예를 들어, 사용자 단말(1100)은 수신된 사용자 입력에 대응되는 플랜을 수신하고, 플랜에 따라 '이번주 일정'을 디스플레이에 표시할 수 있다.
- [0203] 본 문서에 개시된 다양한 실시예들에 따른 전자 장치는 다양한 형태의 장치가 될 수 있다. 전자 장치는, 예를 들면, 휴대용 통신 장치 (예: 스마트폰), 컴퓨터 장치, 휴대용 멀티미디어 장치, 휴대용 의료 기기, 카메라, 웨어러블 장치, 또는 가전 장치를 포함할 수 있다. 본 문서의 실시예에 따른 전자 장치는 전술한 기기들에 한정되

지 않는다.

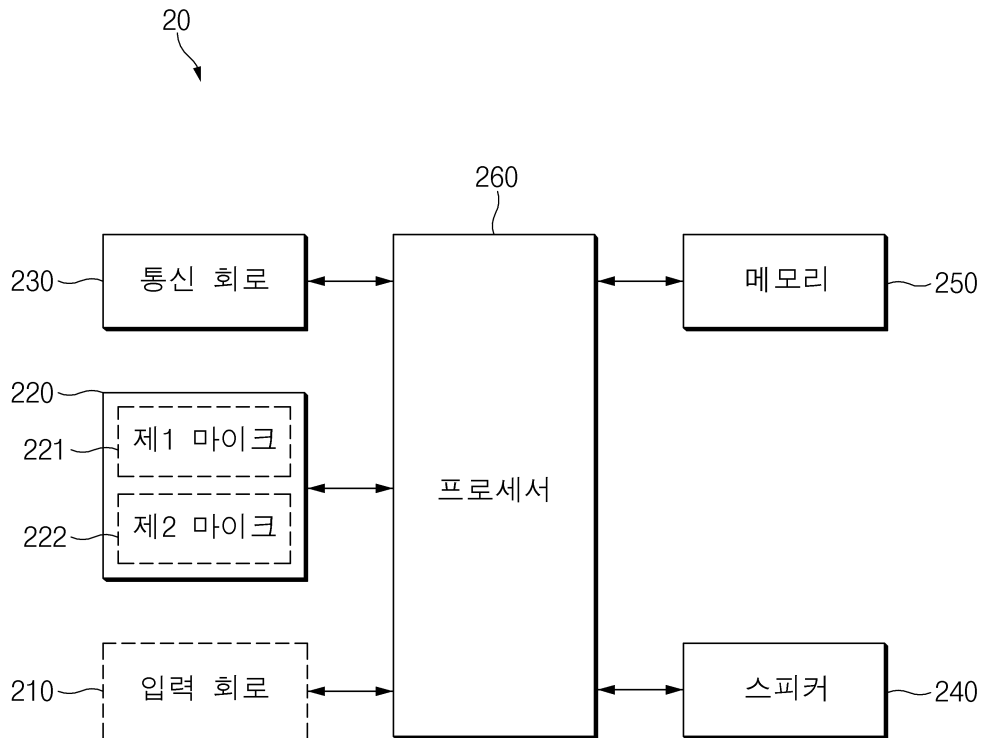
- [0204] 본 문서의 다양한 실시예들 및 이에 사용된 용어들은 본 문서에 기재된 기술적 특징들을 특정한 실시예들로 한정하려는 것이 아니며, 해당 실시예의 다양한 변경, 균등물, 또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 또는 관련된 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 아이টে에 대응하는 명사의 단수 형은 관련된 문맥상 명백하게 다르게 지시하지 않는 한, 상기 아이টে 한 개 또는 복수 개를 포함할 수 있다. 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C", "A, B 및 C 중 적어도 하나," 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들 중 어느 하나, 또는 그들의 모든 가능한 조합을 포함할 수 있다. "제 1", "제 2", 또는 "첫째" 또는 "둘째"와 같은 용어들은 단순히 해당 구성요소를 다른 해당 구성요소와 구분하기 위해 사용될 수 있으며, 해당 구성요소들을 다른 측면(예: 중요성 또는 순서)에서 한정하지 않는다. 어떤(예: 제 1) 구성요소가 다른(예: 제 2) 구성요소에, "기능적으로" 또는 "통신적으로"라는 용어와 함께 또는 이런 용어 없이, "커플드" 또는 "커넥티드"라고 언급된 경우, 그것은 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로(예: 유선으로), 무선으로, 또는 제 3 구성요소를 통하여 연결될 수 있다는 것을 의미한다.
- [0205] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어로 구현된 유닛을 포함할 수 있으며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. 모듈은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는, 상기 부품의 최소 단위 또는 그 일부가 될 수 있다. 예를 들면, 일실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.
- [0206] 본 문서의 다양한 실시예들은 기기(machine)(예: 전자 장치(1001)) 의해 읽을 수 있는 저장 매체(storage medium)(예: 내장 메모리(1036) 또는 외장 메모리(1038))에 저장된 하나 이상의 명령어들을 포함하는 소프트웨어(예: 프로그램(1040))로서 구현될 수 있다. 예를 들면, 기기(예: 전자 장치(1001))의 프로세서(예: 프로세서(1020))는, 저장 매체로부터 저장된 하나 이상의 명령어들 중 적어도 하나의 명령을 호출하고, 그것을 실행할 수 있다. 이것은 기기가 상기 호출된 적어도 하나의 명령어에 따라 적어도 하나의 기능을 수행하도록 운영되는 것을 가능하게 한다. 상기 하나 이상의 명령어들은 컴파일러에 의해 생성된 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다. 기기로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 실재(tangible)하는 장치이고, 신호(signal)(예: 전자기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장매체에 반영구적으로 저장되는 경우와 임시적으로 저장되는 경우를 구분하지 않는다.
- [0207] 일실시예에 따르면, 본 문서에 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 제품의 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.
- [0208] 다양한 실시예들에 따르면, 상기 기술한 구성요소들의 각각의 구성요소(예: 모듈 또는 프로그램)는 단수 또는 복수의 개체를 포함할 수 있다. 다양한 실시예들에 따르면, 전술한 해당 구성요소들 중 하나 이상의 구성요소들 또는 동작들이 생략되거나, 또는 하나 이상의 다른 구성요소들 또는 동작들이 추가될 수 있다. 대체적으로 또는 추가적으로, 복수의 구성요소들(예: 모듈 또는 프로그램)은 하나의 구성요소로 통합될 수 있다. 이런 경우, 통합된 구성요소는 상기 복수의 구성요소들 각각의 구성요소의 하나 이상의 기능들을 상기 통합 이전에 상기 복수의 구성요소들 중 해당 구성요소에 의해 수행되는 것과 동일 또는 유사하게 수행할 수 있다. 다양한 실시예들에 따르면, 모듈, 프로그램 또는 다른 구성요소에 의해 수행되는 동작들은 순차적으로, 병렬적으로, 반복적으로, 또는 휴리스틱하게 실행되거나, 상기 동작들 중 하나 이상이 다른 순서로 실행되거나, 생략되거나, 또는 하나 이상의 다른 동작들이 추가될 수 있다.

도면

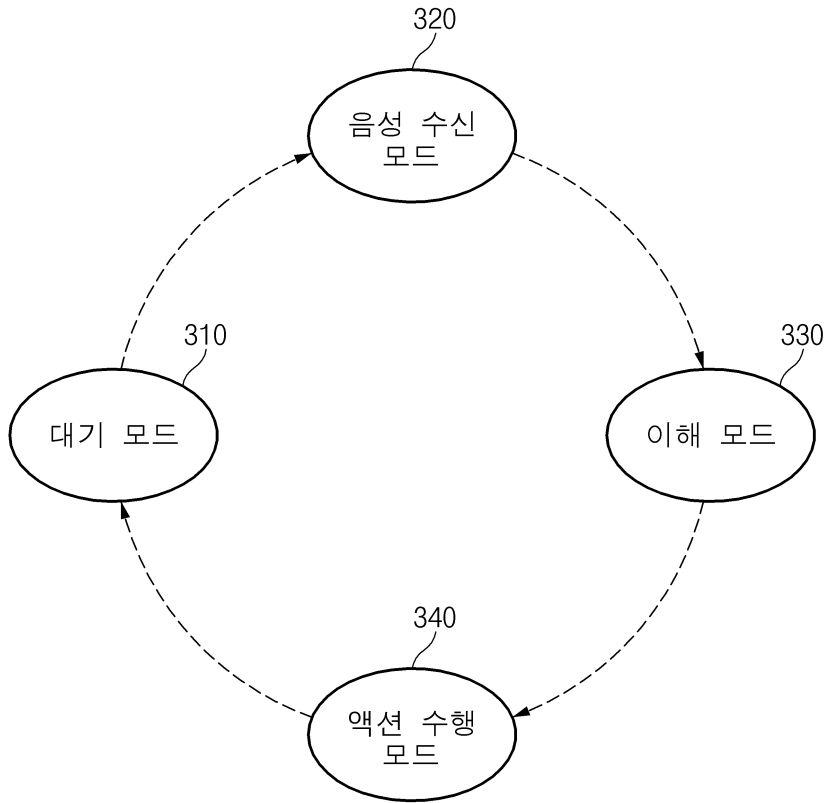
도면1



도면2



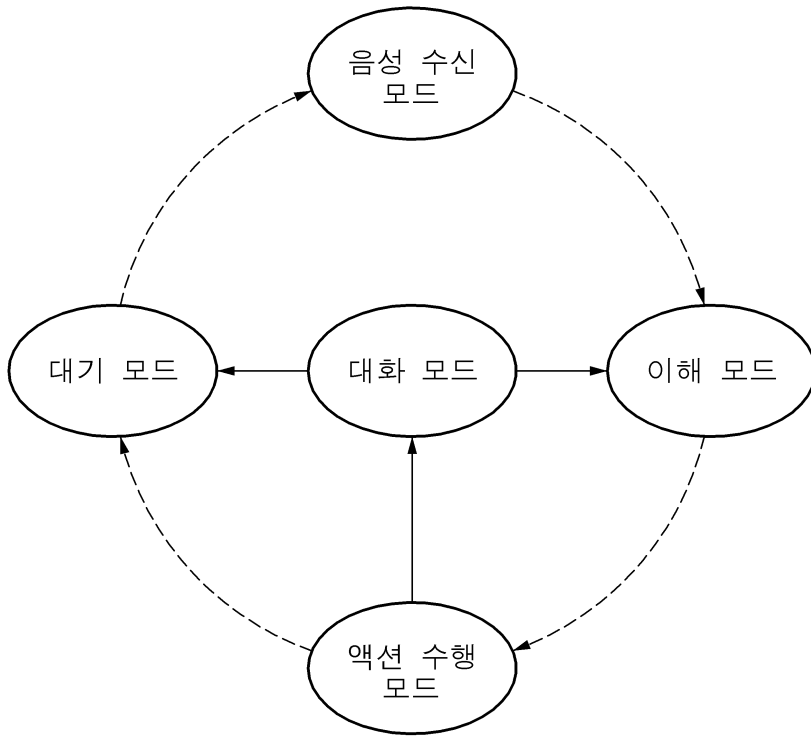
도면3a



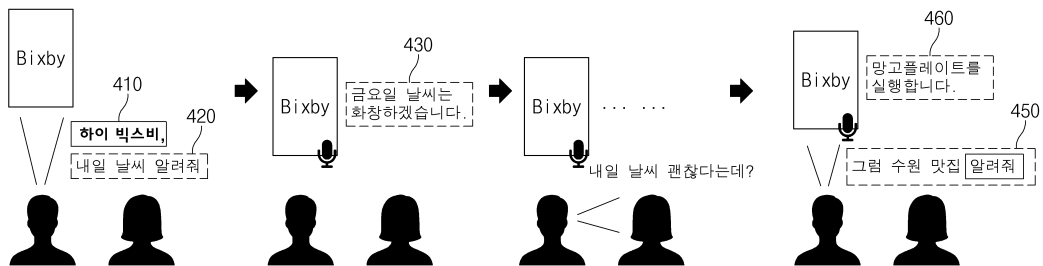
도면3b

사용자: ³⁵¹하이 빅스비, ³⁵²내일 날씨 알려줘
 전자 장치: ³⁵³금요일 날씨는 화창하겠습니다.
 사용자: ³⁵⁴하이 빅스비, ³⁵⁵수원 맛집 알려줘
 전자 장치: ³⁶⁵수원 맛집은 XXX와 XXA가 있습니다.
 사용자: ³⁶⁷하이 빅스비, ³⁶⁸토요일 날씨 알려줘
 전자 장치: ³⁶⁹토요일 날씨는 흐리겠습니다.

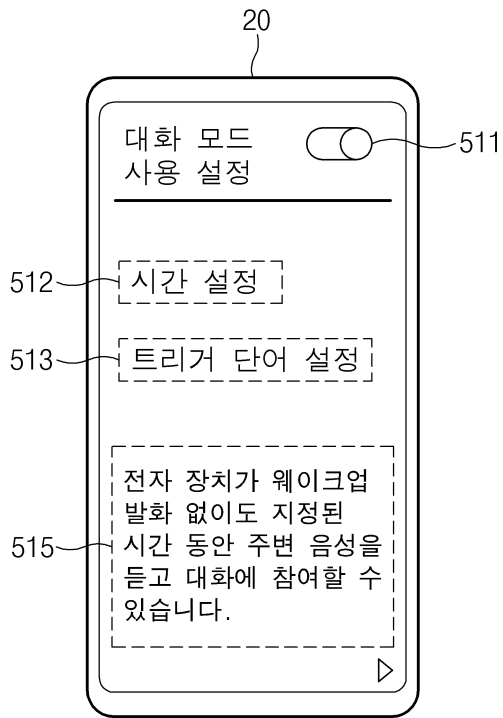
도면4a



도면4b

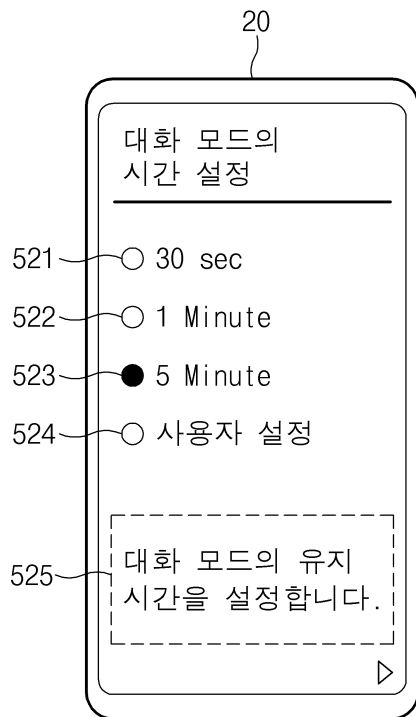


도면5a



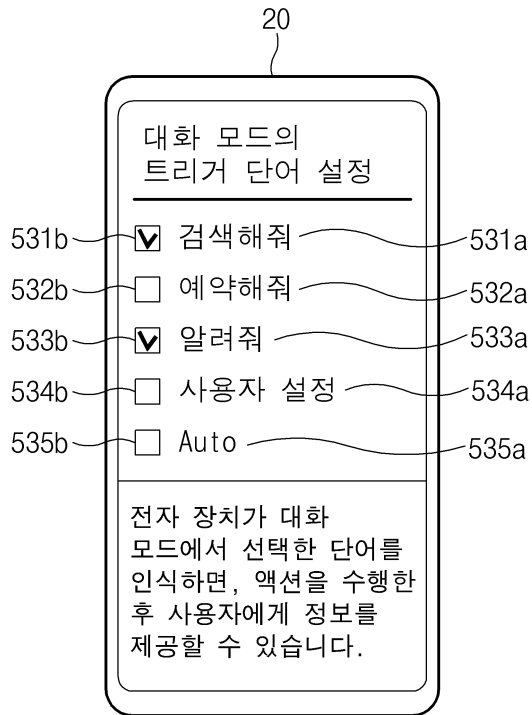
<510>

도면5b



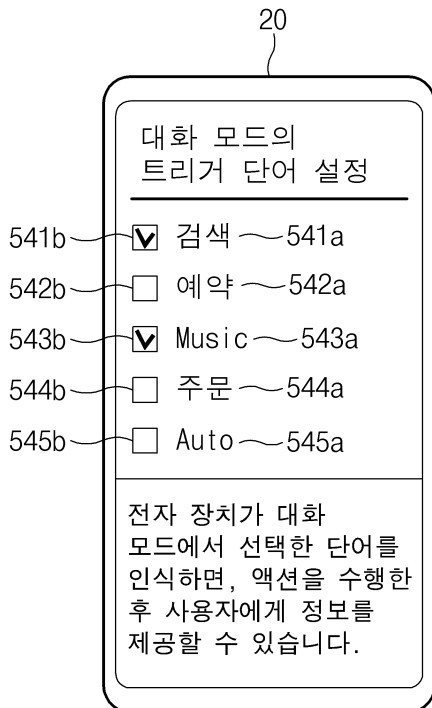
<520>

도면5c



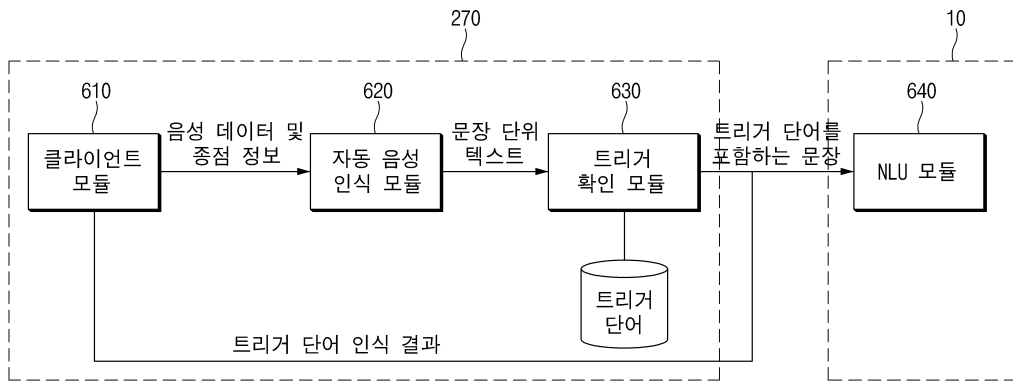
<530>

도면5d

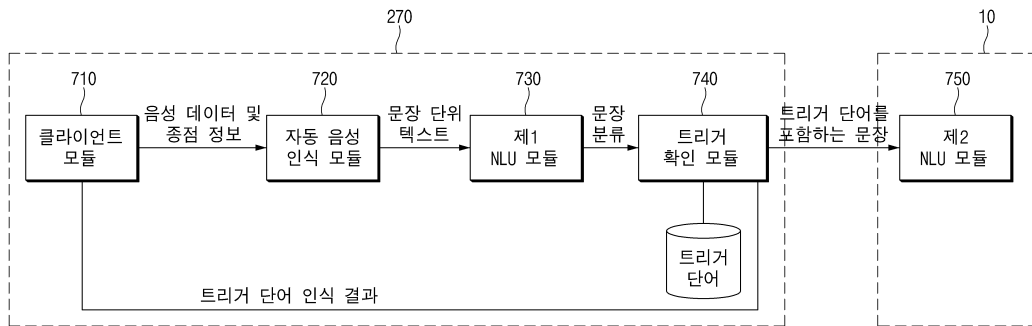


<540>

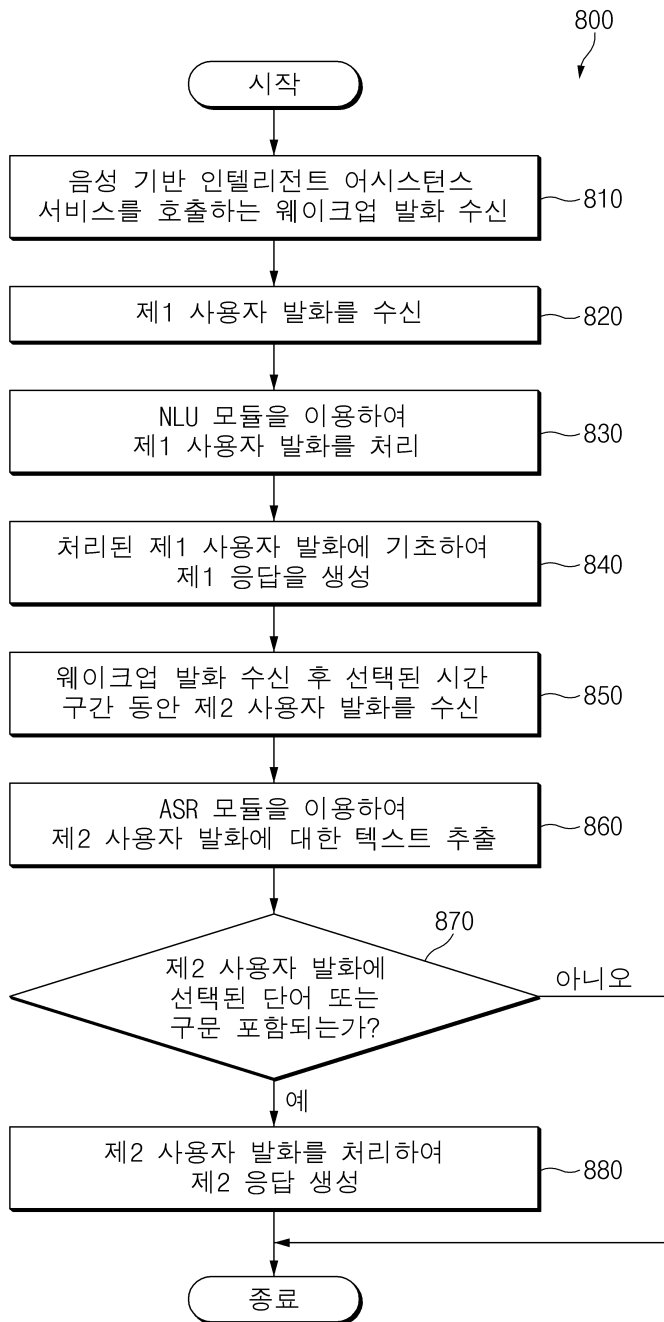
도면6



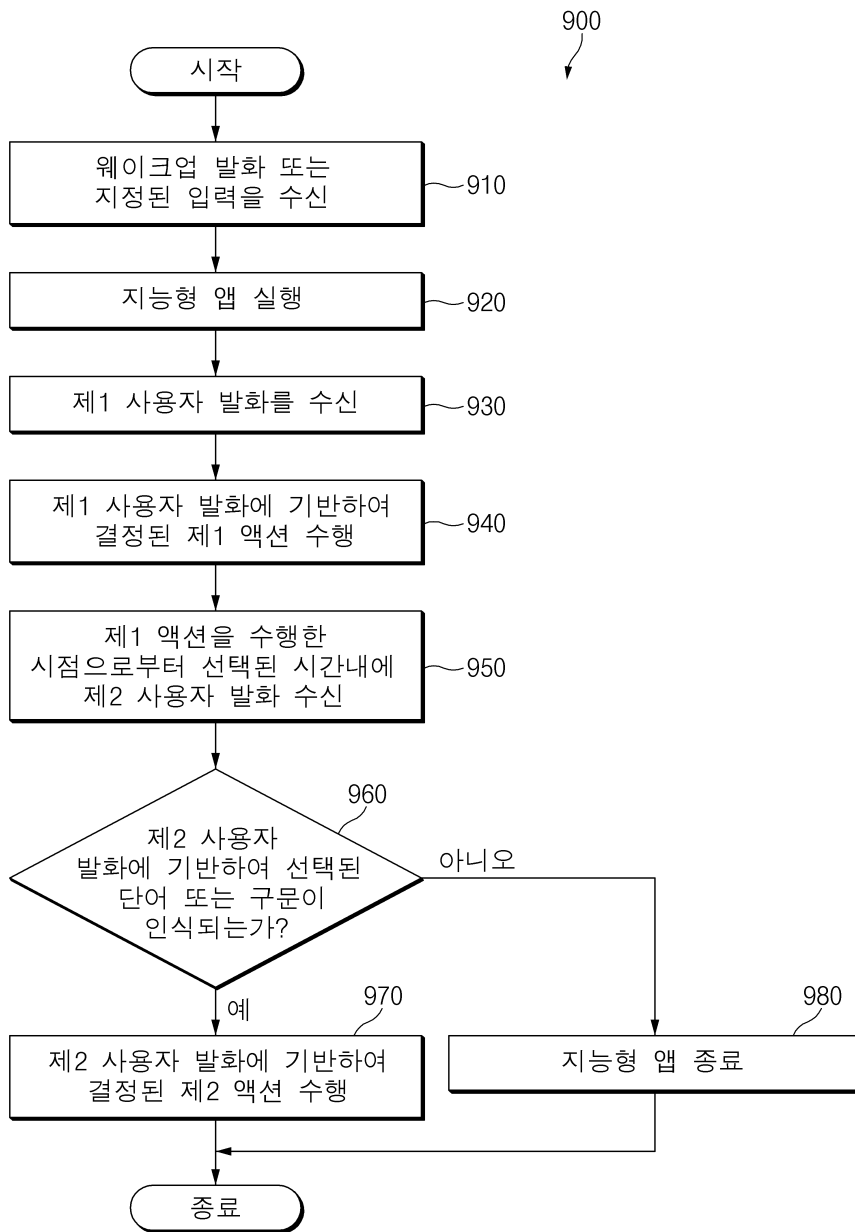
도면7



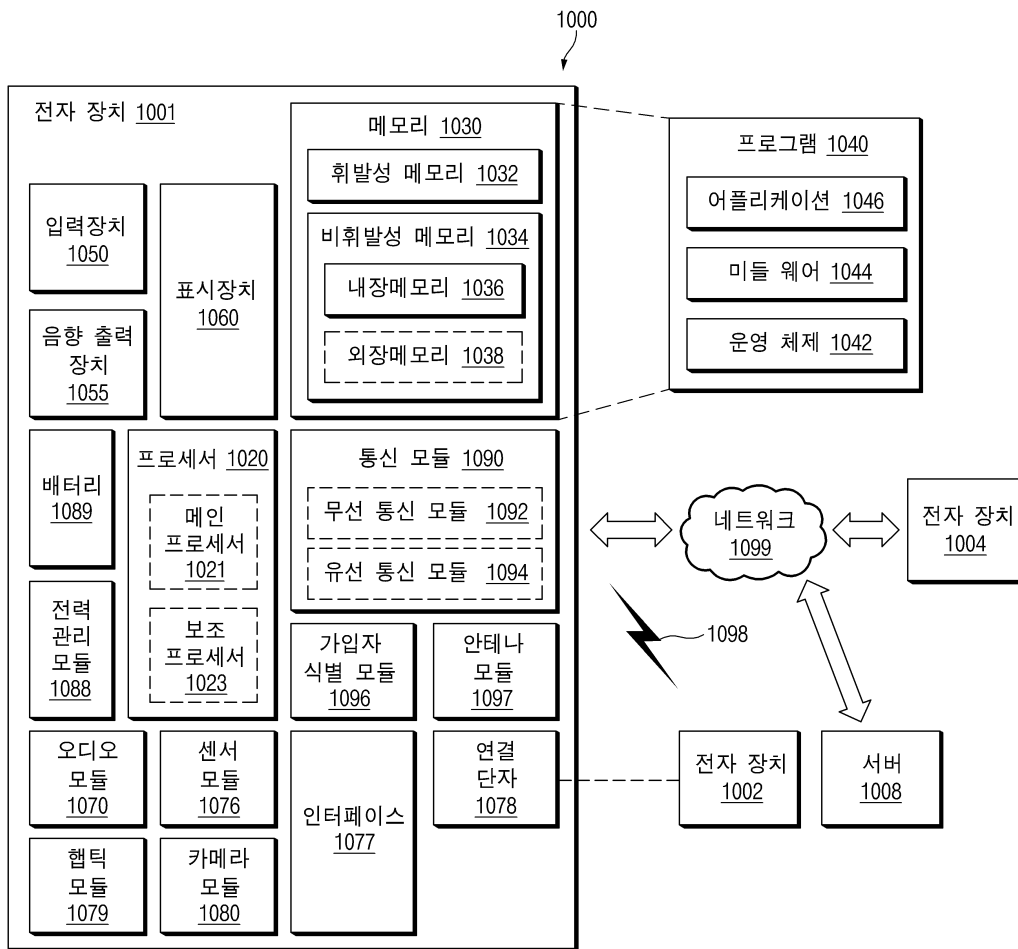
도면8



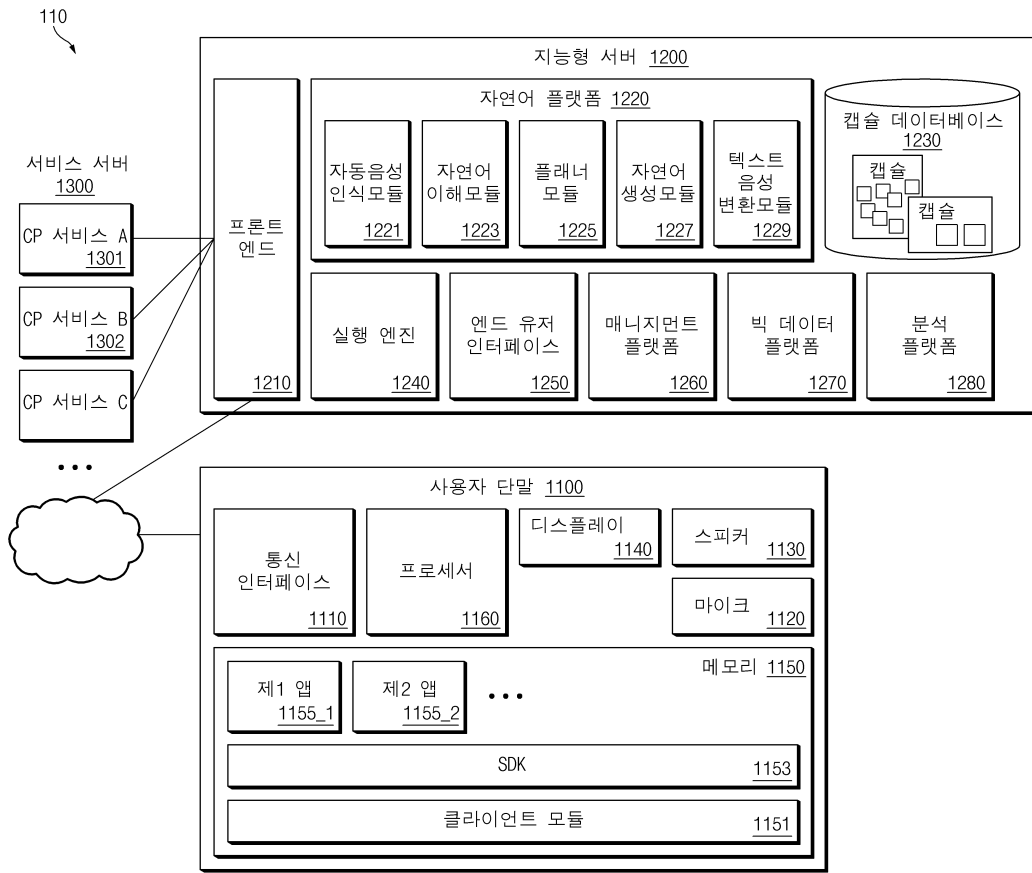
도면9



도면10



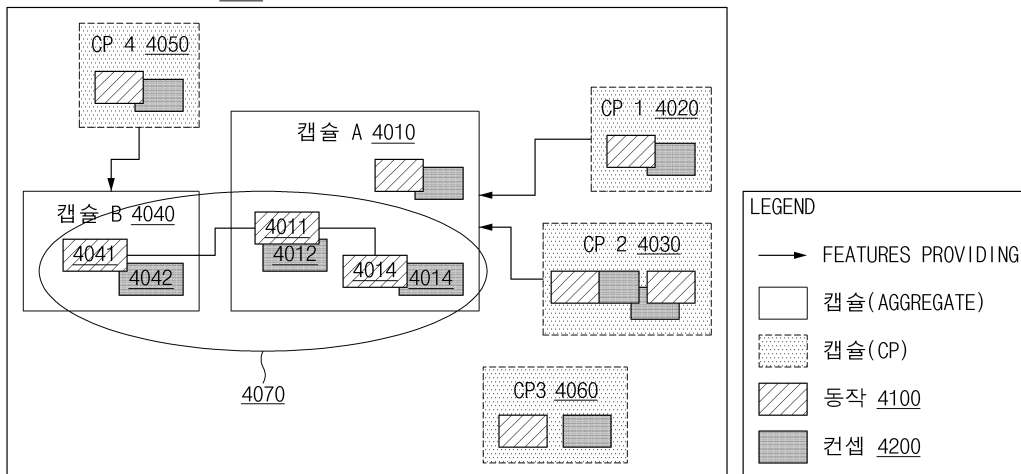
도면11



1155 : 1155_1, 1155_2

도면12

컨셉 액션 네트워크 4000



도면13

