



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0073733
(43) 공개일자 2020년06월24일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
 <i>G06F 3/01</i> (2006.01) <i>G06F 3/00</i> (2006.01)
 <i>G06F 3/16</i> (2018.01) <i>G06K 9/00</i> (2006.01)
 <i>G10L 15/04</i> (2006.01) <i>G10L 15/25</i> (2013.01)</p> <p>(52) CPC특허분류
 <i>G06F 3/011</i> (2013.01)
 <i>G06F 3/005</i> (2013.01)</p> <p>(21) 출원번호 10-2018-0162246
 (22) 출원일자 2018년12월14일
 심사청구일자 없음</p> | <p>(71) 출원인
 삼성전자주식회사
 경기도 수원시 영통구 삼성로 129 (매탄동)</p> <p>(72) 발명자
 김선욱
 서울특별시 서초구 강남대로39길 15-10, 911호(서초동)
 장성운
 경기도 안양시 동안구 관평로 68, 708동 101호(평촌동, 꿈마을한신아파트)
 우혜림
 경기도 수원시 권선구 권광로 55, 105동 702호(권선동, 권선자이 이편한세상)</p> <p>(74) 대리인
 윤엔리특허법인(유한)</p> |
|--|---|

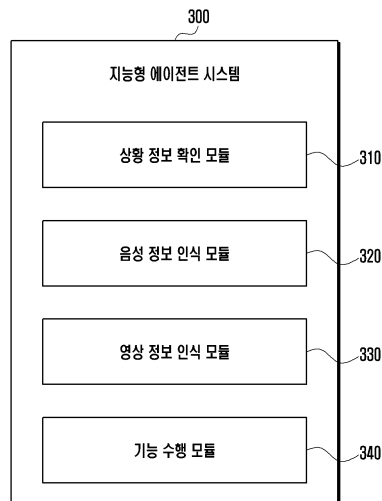
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 전자 장치의 기능 실행 방법 및 이를 사용하는 전자 장치

(57) 요약

전자 장치에 있어서, 카메라; 마이크; 디스플레이; 메모리; 및 프로세서를 포함하고, 상기 프로세서는, 적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하고, 상기 전자 장치의 상황 정보를 확인하고, 상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하도록 제어하고, 상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하고, 및 상기 인식된 발화에 상응하는 기능을 수행하도록 설정될 수 있다. 이 밖의 다른 실시예도 가능하다.

대표도 - 도3



(52) CPC특허분류

G06F 3/167 (2013.01)

G06K 9/00335 (2013.01)

G10L 15/04 (2013.01)

G10L 15/25 (2013.01)

명세서

청구범위

청구항 1

전자 장치에 있어서,

카메라;

마이크;

디스플레이;

메모리; 및

프로세서를 포함하고, 상기 프로세서는,

적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하고,

상기 전자 장치의 상황 정보를 확인하고,

상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하도록 제어하고,

상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하고, 및

상기 인식된 발화에 상응하는 기능을 수행하도록 설정된 전자 장치.

청구항 2

제1 항에 있어서,

상기 프로세서는,

상기 마이크를 통해, 상기 전자 장치의 주변 소음을 획득하고, 및

상기 획득한 주변 소음에 관한 정보를 상기 상황 정보로 저장하도록 설정된 전자 장치.

청구항 3

제2 항에 있어서,

상기 프로세서는,

상기 주변 소음이 기 설정된 값 이상인 경우, 상기 카메라를 활성화 하도록 설정된 전자 장치.

청구항 4

제1 항에 있어서,

상기 프로세서는,

상기 실행 중인 적어도 하나의 어플리케이션의 종류 또는 실행 상태 중 적어도 하나를 상기 상황 정보로 저장하고, 및

상기 실행 중인 적어도 하나의 어플리케이션이 음악 또는 동영상 재생 중인 경우, 상기 카메라를 활성화 하도록 설정된 전자 장치.

청구항 5

제1 항에 있어서,
상기 프로세서는,
상기 마이크를 통해, 음성 신호를 수신하도록 제어하고,
상기 음성 신호 및 상기 입술의 움직임의 조합에 상응하는 조합 발화를 인식하고, 및
상기 인식된 조합 발화에 상응하는 기능을 수행하도록 설정된 전자 장치.

청구항 6

제5 항에 있어서,
상기 프로세서는,
상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 상기 입술의 움직임을 제외한, 상기 음성 신호에 상응하는 음성 인식 기반 발화를 인식하고, 및
상기 인식된 음성 인식 기반 발화에 상응하는 기능을 수행하도록 설정된 전자 장치.

청구항 7

제5 항에 있어서,
상기 프로세서는,
상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 사용자에게 상기 입술의 움직임을 인식하는데 실패하였음을 알리는 사용자 인터페이스를 상기 디스플레이를 통해 표시하도록 설정된 전자 장치.

청구항 8

제5 항에 있어서,
상기 프로세서는,
상기 획득한 영상 정보를 기반으로, 복수 개의 입술을 감지하는 경우, 상기 음성 신호 및 상기 복수 개의 입술의 움직임들을 확인하고, 및
상기 사용자에게 해당하는 입술을 다른 입술과 구분되게 하는 사용자 인터페이스를 상기 디스플레이를 통해 표시하도록 설정된 전자 장치.

청구항 9

제8 항에 있어서,
상기 프로세서는,
상기 사용자에게 해당하는 입술의 움직임을 기반으로 인식된 제1 발화의 가중치를 상기 마이크를 통해 획득한 음성 신호를 기반으로 인식된 제2 발화의 가중치보다 더 높게 설정하여, 상기 조합 발화를 인식하도록 설정된 전자 장치.

청구항 10

제1 항에 있어서,

상기 프로세서는,

상기 지능형 에이전트 서비스가 활성화 된 경우에도, 상기 적어도 하나의 어플리케이션에 관한 적어도 하나의 정보를 끊김 없이(seamless) 제공하도록 설정된 전자 장치.

청구항 11

전자 장치의 방법에 있어서,

적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하는 동작;

상기 전자 장치의 상황 정보를 확인하는 동작;

상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하는 동작;

상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하는 동작; 및

상기 인식된 발화에 상응하는 기능을 수행하는 동작을 포함하는 전자 장치의 방법.

청구항 12

제11 항에 있어서,

마이크를 통해, 상기 전자 장치의 주변 소음을 획득하는 동작; 및

상기 획득한 주변 소음에 관한 정보를 상기 상황 정보로 저장하는 동작을 포함하는 전자 장치의 방법.

청구항 13

제12 항에 있어서,

상기 주변 소음이 기 설정된 값 이상인 경우, 상기 카메라를 활성화 하는 동작을 포함하는 전자 장치의 방법.

청구항 14

제11 항에 있어서,

상기 실행 중인 적어도 하나의 어플리케이션의 종류 또는 실행 상태 중 적어도 하나를 상기 상황 정보로 저장하는 동작; 및

상기 실행 중인 적어도 하나의 어플리케이션이 음악 또는 동영상 재생 중인 경우, 상기 카메라를 활성화 하는 동작을 포함하는 전자 장치의 방법.

청구항 15

제11 항에 있어서,

마이크를 통해, 음성 신호를 수신하는 동작;

상기 음성 신호 및 상기 입술의 움직임을 조합에 상응하는 조합 발화를 인식하는 동작; 및

상기 인식된 조합 발화에 상응하는 기능을 수행하는 동작을 포함하는 전자 장치의 방법.

청구항 16

제15 항에 있어서,

상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 상기 입술의 움직임에 제외한, 상기 음성 신호에 상응하는 음성 인식 기반 발화를 인식하는 동작; 및

상기 인식된 음성 인식 기반 발화에 상응하는 기능을 수행하는 동작을 포함하는 전자 장치의 방법.

청구항 17

제15 항에 있어서,

상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 사용자에게 상기 입술의 움직임을 인식하는데 실패하였음을 알리는 사용자 인터페이스를 디스플레이를 통해 표시하는 동작을 포함하는 전자 장치의 방법.

청구항 18

제15 항에 있어서,

상기 획득한 영상 정보를 기반으로, 복수 개의 입술을 감지하는 경우, 상기 음성 신호 및 상기 복수 개의 입술의 움직임을 확인하는 동작; 및

상기 사용자에게 해당하는 입술을 다른 입술과 구분되게 하는 사용자 인터페이스를 디스플레이를 통해 표시하는 동작을 포함하는 전자 장치의 방법.

청구항 19

제18 항에 있어서,

상기 사용자에게 해당하는 입술의 움직임을 기반으로 인식된 제1 발화의 가중치를 상기 마이크를 통해 획득한 음성 신호를 기반으로 인식된 제2 발화의 가중치보다 더 높게 설정하여, 상기 조합 발화를 인식하는 동작을 포함하는 전자 장치의 방법.

청구항 20

제11 항에 있어서,

상기 지능형 에이전트 서비스가 활성화 된 경우에도, 상기 적어도 하나의 어플리케이션에 관한 적어도 하나의 정보를 끊김 없이(seamless) 제공하는 동작을 포함하는 전자 장치의 방법.

발명의 설명

기술 분야

[0001] 본 발명의 다양한 실시 예는 전자 장치의 기능 실행 방법 및 이를 사용하는 전자 장치에 관한 것이다.

배경 기술

[0002] 오늘 날 스마트폰, 태블릿 PC(personal computer), 웨어러블 디바이스(wearable device) 등과 같은 다양한 종

류의 전자 장치가 널리 사용되고 있다. 전자 장치의 활용을 높이기 위해, 하드웨어 구성뿐만 아니라 소프트웨어 구성 역시 지속적으로 개량되고 있다. 예를 들어, 전자 장치는 사용자 음성 입력에 응답하여 다양한 기능을 수행하는 지능형 에이전트(intelligence agent) 서비스를 제공할 수 있다.

발명의 내용

해결하려는 과제

[0003] 지능형 에이전트 서비스는 음성을 인식하고, 인식된 음성을 분석하여, 사용자가 수행하고자 하는 서비스를 사용자에게 제공할 수 있다. 이러한 지능형 에이전트 서비스는 사용자의 음성을 기반으로 동작하는 것이기 때문에, 사용자 음성 이외의 노이즈(예: 주변 소음)에 따라 음성 인식 성능이 영향 받을 수 있다. 예를 들어, 전자 장치 주변이 조용한 상태(예: 소음이 적은 상태)에서 사용자 음성을 인식하는 것과 전자 장치 주변이 시끄러운 상태(예: 소음이 많은 상태)에서 사용자 음성을 인식하는 것은 상당한 성능 차이가 발생할 수 있다.

과제의 해결 수단

[0004] 다양한 실시 예에 따른 전자 장치에 있어서, 카메라; 마이크; 디스플레이; 메모리; 및 프로세서를 포함하고, 상기 프로세서는, 적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하고, 상기 전자 장치의 상황 정보를 확인하고, 상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하도록 제어하고, 상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하고, 및 상기 인식된 발화에 상응하는 기능을 수행하도록 설정될 수 있다.

[0005] 다양한 실시 예에 따른 전자 장치의 방법에 있어서, 적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하는 동작; 상기 전자 장치의 상황 정보를 확인하는 동작; 상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하는 동작; 상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하는 동작; 및 상기 인식된 발화에 상응하는 기능을 수행하는 동작을 포함할 수 있다.

발명의 효과

[0006] 본 발명의 다양한 실시 예에 따른 전자 장치는 지능형 에이전트 서비스의 정확도를 향상시키기 위해 영상 정보를 활용할 수 있다. 예를 들어, 전자 장치는 획득한 영상을 기반으로 립 리딩 기술을 적용하여 사용자가 수행하려는 기능을 분석할 수 있다. 전자 장치는 사용자의 음성 또는 사용자의 입술 모양 중 적어도 하나를 기반으로 사용자가 원하는 기능을 결정할 수 있으며, 결정된 기능을 사용자에게 제공할 수 있다. 전자 장치는 사용자의 입술 모양을 기반으로 사용자 발화가 시작되는 시점과 사용자 발화가 종료되는 시점을 검출할 수 있으며, 불명확하게 수신된 사용자 음성을 입술 모양을 통해 교정할 수 있으므로 사용자에게 좀 더 정확한 지능형 에이전트 서비스를 제공할 수 있다.

도면의 간단한 설명

[0007] 도 1은 본 발명의 다양한 실시 예에 따른 네트워크 환경 내의 전자 장치의 블록도이다.
 도 2A는 일 실시예에 따른 통합 지능(integrated intelligence) 시스템을 나타낸 블록도이다.
 도 2B는 일 실시예에 따른, 컨셉과 액션의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.
 도 2C는 일 실시예에 따라, 지능형 앱을 통해 수신된 음성 입력을 처리하는 화면을 표시하는 사용자 단말을 도시하는 도면이다.
 도 3은 본 발명의 다양한 실시 예에 따른 지능형 에이전트 시스템을 개략적으로 나타낸 도면이다.
 도 4는 본 발명의 다양한 실시 예에 따른 립 리딩 모듈의 블록도이다.
 도 5는 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 사용자 명령에 상응하는 기능을 수행하는 방법에 관한 흐름도이다.
 도 6은 본 발명의 다양한 실시 예에 따른 전자 장치에서 사용자의 음성 또는 사용자의 입술 모양 중 적어도 하나에 기반하여, 사용자 발화를 인식하는 방법을 나타내는 도면이다.

도 7은 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 기능을 수행하는 경우의 사용자 인터페이스를 나타내는 도면이다.

도 8A 내지 8B는 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 기능을 수행하는 경우의 사용자 인터페이스를 나타내는 도면이다.

도 9는 본 발명의 다양한 실시 예에 따른 전자 장치의 립 리딩이 가능한지 여부에 기반하여 사용자 명령에 상응하는 기능을 수행하는 방법에 관한 흐름도이다.

도 10A 내지 10C는 본 발명의 다양한 실시 예에 따른 전자 장치의 립 리딩 동작에 관한 사용자 인터페이스를 제공하는 방법을 개략적으로 나타낸 도면이다.

도 11은 본 발명의 다양한 실시 예에 따른 어플리케이션 실행 중 입술의 움직임을 감지하여 사용자의 발화를 인식하고, 인식된 발화에 상응하는 기능을 수행하는 방법에 관한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0008] 도 1은, 다양한 실시예들에 따른, 네트워크 환경(100) 내의 전자 장치(101)의 블록도이다. 도 1을 참조하면, 네트워크 환경(100)에서 전자 장치(101)는 제 1 네트워크(198)(예: 근거리 무선 통신 네트워크)를 통하여 전자 장치(102)와 통신하거나, 또는 제 2 네트워크(199)(예: 원거리 무선 통신 네트워크)를 통하여 전자 장치(104) 또는 서버(108)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(101)는 서버(108)를 통하여 전자 장치(104)와 통신할 수 있다. 일실시예에 따르면, 전자 장치(101)는 프로세서(120), 메모리(130), 입력 장치(150), 음향 출력 장치(155), 표시 장치(160), 오디오 모듈(170), 센서 모듈(176), 인터페이스(177), 햅틱 모듈(179), 카메라 모듈(180), 전력 관리 모듈(188), 배터리(189), 통신 모듈(190), 가입자 식별 모듈(196), 또는 안테나 모듈(197)을 포함할 수 있다. 어떤 실시예에서는, 전자 장치(101)에는, 이 구성요소들 중 적어도 하나(예: 표시 장치(160) 또는 카메라 모듈(180))가 생략되거나, 하나 이상의 다른 구성 요소가 추가될 수 있다. 어떤 실시예에서는, 이 구성요소들 중 일부들은 하나의 통합된 회로로 구현될 수 있다. 예를 들면, 센서 모듈(176)(예: 지문 센서, 홍채 센서, 또는 조도 센서)은 표시 장치(160)(예: 디스플레이)에 임베디드된 채 구현될 수 있다.
- [0009] 프로세서(120)는, 예를 들면, 소프트웨어(예: 프로그램(140))를 실행하여 프로세서(120)에 연결된 전자 장치(101)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(120)는 다른 구성요소(예: 센서 모듈(176) 또는 통신 모듈(190))로부터 수신된 명령 또는 데이터를 휘발성 메모리(132)에 로드하고, 휘발성 메모리(132)에 저장된 명령 또는 데이터를 처리하고, 결과 데이터를 비휘발성 메모리(134)에 저장할 수 있다. 일실시예에 따르면, 프로세서(120)는 메인 프로세서(121)(예: 중앙 처리 장치 또는 어플리케이션 프로세서), 및 이와는 독립적으로 또는 함께 운영 가능한 보조 프로세서(123)(예: 그래픽 처리 장치, 이미지 시그널 프로세서, 센서 허브 프로세서, 또는 커뮤니케이션 프로세서)를 포함할 수 있다. 추가적으로 또는 대체적으로, 보조 프로세서(123)는 메인 프로세서(121)보다 저전력을 사용하거나, 또는 지정된 기능에 특화되도록 설정될 수 있다. 보조 프로세서(123)는 메인 프로세서(121)와 별개로, 또는 그 일부로서 구현될 수 있다.
- [0010] 보조 프로세서(123)는, 예를 들면, 메인 프로세서(121)가 인액티브(예: 슬립) 상태에 있는 동안 메인 프로세서(121)를 대신하여, 또는 메인 프로세서(121)가 액티브(예: 어플리케이션 실행) 상태에 있는 동안 메인 프로세서(121)와 함께, 전자 장치(101)의 구성요소들 중 적어도 하나의 구성요소(예: 표시 장치(160), 센서 모듈(176), 또는 통신 모듈(190))와 관련된 기능 또는 상태들의 적어도 일부를 제어할 수 있다. 일실시예에 따르면, 보조 프로세서(123)(예: 이미지 시그널 프로세서 또는 커뮤니케이션 프로세서)는 기능적으로 관련 있는 다른 구성 요소(예: 카메라 모듈(180) 또는 통신 모듈(190))의 일부로서 구현될 수 있다.
- [0011] 메모리(130)는, 전자 장치(101)의 적어도 하나의 구성요소(예: 프로세서(120) 또는 센서모듈(176))에 의해 사용되는 다양한 데이터를 저장할 수 있다. 데이터는, 예를 들어, 소프트웨어(예: 프로그램(140)) 및, 이와 관련된 명령에 대한 입력 데이터 또는 출력 데이터를 포함할 수 있다. 메모리(130)는, 휘발성 메모리(132) 또는 비휘발성 메모리(134)를 포함할 수 있다.
- [0012] 프로그램(140)은 메모리(130)에 소프트웨어로서 저장될 수 있으며, 예를 들면, 운영 체제(142), 미들 웨어(144) 또는 어플리케이션(146)을 포함할 수 있다.
- [0013] 입력 장치(150)는, 전자 장치(101)의 구성요소(예: 프로세서(120))에 사용될 명령 또는 데이터를 전자 장치

(101)의 외부(예: 사용자)로부터 수신할 수 있다. 입력 장치(150)은, 예를 들면, 마이크, 마우스, 키보드, 또는 디지털 펜(예: 스타일러스 펜)을 포함할 수 있다.

- [0014] 음향 출력 장치(155)는 음향 신호를 전자 장치(101)의 외부로 출력할 수 있다. 음향 출력 장치(155)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.
- [0015] 표시 장치(160)는 전자 장치(101)의 외부(예: 사용자)로 정보를 시각적으로 제공할 수 있다. 표시 장치(160)은, 예를 들면, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 일실시예에 따르면, 표시 장치(160)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.
- [0016] 오디오 모듈(170)은 소리를 전기 신호로 변환시키거나, 반대로 전기 신호를 소리로 변환시킬 수 있다. 일실시예에 따르면, 오디오 모듈(170)은, 입력 장치(150)를 통해 소리를 획득하거나, 음향 출력 장치(155), 또는 전자 장치(101)와 직접 또는 무선으로 연결된 외부 전자 장치(예: 전자 장치(102)) (예: 스피커 또는 헤드폰))를 통해 소리를 출력할 수 있다.
- [0017] 센서 모듈(176)은 전자 장치(101)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(176)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.
- [0018] 인터페이스(177)는 전자 장치(101)이 외부 전자 장치(예: 전자 장치(102))와 직접 또는 무선으로 연결되기 위해 사용될 수 있는 하나 이상의 지정된 프로토콜들을 지원할 수 있다. 일실시예에 따르면, 인터페이스(177)는, 예를 들면, HDMI(high definition multimedia interface), USB(universal serial bus) 인터페이스, SD카드 인터페이스, 또는 오디오 인터페이스를 포함할 수 있다.
- [0019] 연결 단자(178)는, 그를 통해서 전자 장치(101)가 외부 전자 장치(예: 전자 장치(102))와 물리적으로 연결될 수 있는 커넥터를 포함할 수 있다. 일실시예에 따르면, 연결 단자(178)는, 예를 들면, HDMI 커넥터, USB 커넥터, SD 카드 커넥터, 또는 오디오 커넥터(예: 헤드폰 커넥터)를 포함할 수 있다.
- [0020] 햅틱 모듈(179)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(179)은, 예를 들면, 모터, 압전 소자, 또는 전기 자극 장치를 포함할 수 있다.
- [0021] 카메라 모듈(180)은 정지 영상 및 동영상을 촬영할 수 있다. 일실시예에 따르면, 카메라 모듈(180)은 하나 이상의 렌즈들, 이미지 센서들, 이미지 시그널 프로세서들, 또는 플래시들을 포함할 수 있다.
- [0022] 전력 관리 모듈(188)은 전자 장치(101)에 공급되는 전력을 관리할 수 있다. 일실시예에 따르면, 전력 관리 모듈(388)은, 예를 들면, PMIC(power management integrated circuit)의 적어도 일부로서 구현될 수 있다.
- [0023] 배터리(189)는 전자 장치(101)의 적어도 하나의 구성 요소에 전력을 공급할 수 있다. 일실시예에 따르면, 배터리(189)는, 예를 들면, 재충전 불가능한 1차 전지, 재충전 가능한 2차 전지 또는 연료 전지를 포함할 수 있다.
- [0024] 통신 모듈(190)은 전자 장치(101)와 외부 전자 장치(예: 전자 장치(102), 전자 장치(104), 또는 서버(108))간의 직접(예: 유선) 통신 채널 또는 무선 통신 채널의 수립, 및 수립된 통신 채널을 통한 통신 수행을 지원할 수 있다. 통신 모듈(190)은 프로세서(120)(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(190)은 무선 통신 모듈(192)(예: 셀룰러 통신 모듈, 근거리 무선 통신 모듈, 또는 GNSS(global navigation satellite system) 통신 모듈) 또는 유선 통신 모듈(194)(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(198)(예: 블루투스, WiFi direct 또는 IrDA(infrared data association) 같은 근거리 통신 네트워크) 또는 제 2 네트워크(199)(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(192)은 가입자 식별 모듈(196)에 저장된 가입자 정보(예: 국제 모바일 가입자 식별자(IMS))를 이용하여 제 1 네트워크(198) 또

는 제 2 네트워크(199)와 같은 통신 네트워크 내에서 전자 장치(101)를 확인 및 인증할 수 있다.

- [0025] 안테나 모듈(197)은 신호 또는 전력을 외부(예: 외부 전자 장치)로 송신하거나 외부로부터 수신할 수 있다. 일 실시예에 따르면, 안테나 모듈은 서브스트레이트(예: PCB) 위에 형성된 도전체 또는 도전성 패턴으로 이루어진 방사체를 포함하는 하나의 안테나를 포함할 수 있다. 일 실시예에 따르면, 안테나 모듈(197)은 복수의 안테나들을 포함할 수 있다. 이런 경우, 제 1 네트워크(198) 또는 제 2 네트워크(199)와 같은 통신 네트워크에서 사용되는 통신 방식에 적합한 적어도 하나의 안테나가, 예를 들면, 통신 모듈(190)에 의하여 상기 복수의 안테나들로부터 선택될 수 있다. 신호 또는 전력은 상기 선택된 적어도 하나의 안테나를 통하여 통신 모듈(190)과 외부 전자 장치 간에 송신되거나 수신될 수 있다. 어떤 실시예에 따르면, 방사체 이외에 다른 부품(예: RFIC)이 추가로 안테나 모듈(197)의 일부로 형성될 수 있다.
- [0026] 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.
- [0027] 일 실시예에 따르면, 명령 또는 데이터는 제 2 네트워크(199)에 연결된 서버(108)를 통해서 전자 장치(101)와 외부의 전자 장치(104)간에 송신 또는 수신될 수 있다. 전자 장치(102, 104) 각각은 전자 장치(101)와 동일한 또는 다른 종류의 장치일 수 있다. 일 실시예에 따르면, 전자 장치(101)에서 실행되는 동작들의 전부 또는 일부는 외부 전자 장치들(102, 104, or 108) 중 하나 이상의 외부 장치들에서 실행될 수 있다. 예를 들면, 전자 장치(101)가 어떤 기능이나 서비스를 자동으로, 또는 사용자 또는 다른 장치로부터의 요청에 반응하여 수행해야 할 경우에, 전자 장치(101)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(101)로 전달할 수 있다. 전자 장치(101)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.
- [0028] 도 2A는 일 실시예에 따른 통합 지능 (integrated intelligence) 시스템을 나타낸 블록도이다.
- [0029] 도 2A를 참조하면, 일 실시예의 통합 지능화 시스템은 사용자 단말(220), 지능형 서버(200), 및 서비스 서버(230)를 포함할 수 있다.
- [0030] 일 실시 예의 사용자 단말(220)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.
- [0031] 도시된 실시 예에 따르면, 사용자 단말(220)은 통신 인터페이스(221), 마이크(222), 스피커(223), 디스플레이(224), 메모리(226), 또는 프로세서(225)를 포함할 수 있다. 상기 열거된 구성요소들은 서로 작동적으로 또는 전기적으로 연결될 수 있다.
- [0032] 일 실시 예의 통신 인터페이스(221)는 외부 장치와 연결되어 데이터를 송수신하도록 구성될 수 있다. 일 실시 예의 마이크(222)는 소리(예: 사용자 발화)를 수신하여, 전기적 신호로 변환할 수 있다. 일 실시예의 스피커(223)는 전기적 신호를 소리(예: 음성)으로 출력할 수 있다. 일 실시 예의 디스플레이(224)는 이미지 또는 비디오를 표시하도록 구성될 수 있다. 일 실시 예의 디스플레이(224)는 또한 실행되는 앱(app)(또는, 어플리케이션 프로그램(application program))의 그래픽 사용자 인터페이스(graphic user interface)(GUI)를 표시할 수 있다.
- [0033] 일 실시 예의 메모리(226)는 클라이언트 모듈(227), SDK(software development kit)(228), 및 복수의 앱들(229)을 저장할 수 있다. 상기 클라이언트 모듈(227), 및 SDK(228)는 범용적인 기능을 수행하기 위한 프레임워크(framework)(또는, 솔루션 프로그램)를 구성할 수 있다. 또한, 클라이언트 모듈(227) 또는 SDK(228)는 음성 입력을 처리하기 위한 프레임워크를 구성할 수 있다.
- [0034] 일 실시 예의 메모리(226)는 상기 복수의 앱들(229)은 지정된 기능을 수행하기 위한 프로그램일 수 있다. 일 실시 예에 따르면, 복수의 앱들(229)은 제1 앱(229_1), 제2 앱(229_2) 을 포함할 수 있다. 일 실시 예에 따르면, 복수의 앱들(229) 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(229)은

프로세서(225)에 의해 실행되어 상기 복수의 동작들 중 적어도 일부를 순차적으로 실행할 수 있다.

- [0035] 일 실시 예의 프로세서(225)는 사용자 단말(220)의 전반적인 동작을 제어할 수 있다. 예를 들어, 프로세서(225)는 통신 인터페이스(221), 마이크(222), 스피커(223), 및 디스플레이(224)와 전기적으로 연결되어 연결되어 지정된 동작을 수행할 수 있다.
- [0036] 일 실시 예의 프로세서(225)는 또한 상기 메모리(226)에 저장된 프로그램을 실행시켜 지정된 기능을 수행할 수 있다. 예를 들어, 프로세서(225)는 클라이언트 모듈(227) 또는 SDK(228) 중 적어도 하나를 실행하여, 음성 입력을 처리하기 위한 이하의 동작을 수행할 수 있다. 프로세서(225)는, 예를 들어, SDK(228)를 통해 복수의 애플(229)의 동작을 제어할 수 있다. 클라이언트 모듈(227) 또는 SDK(228)의 동작으로 설명된 이하의 동작은 프로세서(225)의 실행에 의한 동작일 수 있다.
- [0037] 일 실시 예의 클라이언트 모듈(227)은 음성 입력을 수신할 수 있다. 예를 들어, 클라이언트 모듈(227)은 마이크(222)를 통해 감지된 사용자 발화에 대응되는 음성 신호를 수신할 수 있다. 상기 클라이언트 모듈(227)은 수신된 음성 입력을 지능형 서버(200)로 송신할 수 있다. 클라이언트 모듈(227)은 수신된 음성 입력과 함께, 사용자 단말(220)의 상태 정보를 지능형 서버(200)로 송신할 수 있다. 상기 상태 정보는, 예를 들어, 애플의 실행 상태 정보일 수 있다.
- [0038] 일 실시 예의 클라이언트 모듈(227)은 수신된 음성 입력에 대응되는 결과를 수신할 수 있다. 예를 들어, 클라이언트 모듈(227)은 지능형 서버(200)에서 상기 수신된 음성 입력에 대응되는 결과를 산출할 수 있는 경우, 수신된 음성 입력에 대응되는 결과를 수신할 수 있다. 클라이언트 모듈(227)은 상기 수신된 결과를 디스플레이(224)에 표시할 수 있다.
- [0039] 일 실시 예의 클라이언트 모듈(227)은 수신된 음성 입력에 대응되는 플랜을 수신할 수 있다. 클라이언트 모듈(227)은 플랜에 따라 애플의 복수의 동작을 실행한 결과를 디스플레이(224)에 표시할 수 있다. 클라이언트 모듈(227)은, 예를 들어, 복수의 동작의 실행 결과를 순차적으로 디스플레이에 표시할 수 있다. 사용자 단말(220)은, 다른 예를 들어, 복수의 동작을 실행한 일부 결과(예: 마지막 동작의 결과)만을 디스플레이에 표시할 수 있다.
- [0040] 일 실시 예에 따르면, 클라이언트 모듈(227)은 지능형 서버(200)로부터 음성 입력에 대응되는 결과를 산출하기 위해 필요한 정보를 획득하기 위한 요청을 수신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(227)은 상기 요청에 대응하여 상기 필요한 정보를 지능형 서버(200)로 송신할 수 있다.
- [0041] 일 실시 예의 클라이언트 모듈(227)은 플랜에 따라 복수의 동작을 실행한 결과 정보를 지능형 서버(200)로 송신할 수 있다. 지능형 서버(200)는 상기 결과 정보를 이용하여 수신된 음성 입력이 올바르게 처리된 것을 확인할 수 있다.
- [0042] 일 실시 예의 클라이언트 모듈(227)은 음성 인식 모듈을 포함할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(227)은 상기 음성 인식 모듈을 통해 제한된 기능을 수행하는 음성 입력을 인식할 수 있다. 예를 들어, 클라이언트 모듈(227)은 지정된 입력(예: 웨이크 업!)을 통해 유익적인 동작을 수행하기 위한 음성 입력을 처리하기 위한 지능형 애플을 수행할 수 있다.
- [0043] 일 실시 예의 지능형 서버(200)는 통신 망을 통해 사용자 단말(220)로부터 사용자 음성 입력과 관련된 정보를 수신할 수 있다. 일 실시 예에 따르면, 지능형 서버(200)는 수신된 음성 입력과 관련된 데이터를 텍스트 데이터(text data)로 변경할 수 있다. 일 실시 예에 따르면, 지능형 서버(200)는 상기 텍스트 데이터에 기초하여 사용자 음성 입력과 대응되는 태스크(task)를 수행하기 위한 플랜(plan)을 생성할 수 있다.
- [0044] 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.
- [0045] 일 실시 예의 지능형 서버(200)는 생성된 플랜에 따른 결과를 사용자 단말(220)로 송신하거나, 생성된 플랜을 사용자 단말(220)로 송신할 수 있다. 일 실시 예에 따르면, 사용자 단말(220)은 플랜에 따른 결과를 디스플레이에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(220)은 플랜에 따른 동작을 실행한 결과를 디스플레이에

표시할 수 있다.

- [0046] 일 실시 예의 지능형 서버(200)는 프론트 엔드(front end)(201), 자연어 플랫폼(natural language platform)(202), 캡슐 데이터베이스(capsule DB)(203), 실행 엔진(execution engine)(209), 엔드 유저 인터페이스(end user interface)(210), 매니지먼트 플랫폼(management platform)(211), 빅 데이터 플랫폼(big data platform)(212), 또는 분석 플랫폼(analytic platform)(213)을 포함할 수 있다.
- [0047] 일 실시 예의 프론트 엔드(201)는 사용자 단말(220)로부터 수신된 음성 입력을 수신할 수 있다. 프론트 엔드(201)는 상기 음성 입력에 대응되는 응답을 송신할 수 있다.
- [0048] 일 실시 예에 따르면, 자연어 플랫폼(202)은 자동 음성 인식 모듈(automatic speech recognition module)(ASR module)(203), 자연어 이해 모듈(natural language understanding module)(NLU module)(204), 플래너 모듈(planner module)(205), 자연어 생성 모듈(natural language generator module)(NLG module)(206) 또는 텍스트 음성 변환 모듈(text to speech module)(TTS module)(207)를 포함할 수 있다.
- [0049] 일 실시 예의 자동 음성 인식 모듈(203)은 사용자 단말(220)로부터 수신된 음성 입력을 텍스트 데이터로 변환할 수 있다. 일 실시 예의 자연어 이해 모듈(204)은 음성 입력의 텍스트 데이터를 이용하여 사용자의 의도를 파악할 수 있다. 예를 들어, 자연어 이해 모듈(204)은 문법적 분석(syntactic analyze) 또는 의미적 분석(semantic analyze)을 수행하여 사용자의 의도를 파악할 수 있다. 일 실시 예의 자연어 이해 모듈(204)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.
- [0050] 일 실시 예의 플래너 모듈(205)은 자연어 이해 모듈(204)에서 결정된 의도 및 파라미터를 이용하여 플랜을 생성할 수 있다. 일 실시 예에 따르면, 플래너 모듈(205)은 상기 결정된 의도에 기초하여 태스크를 수행하기 위해 필요한 복수의 도메인을 결정할 수 있다. 플래너 모듈(205)은 상기 의도에 기초하여 결정된 복수의 도메인 각각에 포함된 복수의 동작을 결정할 수 있다. 일 실시 예에 따르면, 플래너 모듈(205)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(205)은 상기 복수의 동작, 및 상기 복수의 컨셉 사이의 관계를 단계적(또는, 계층적)으로 결정할 수 있다. 예를 들어, 플래너 모듈(205)은 복수의 컨셉에 기초하여 사용자의 의도에 기초하여 결정된 복수의 동작의 실행 순서를 결정할 수 있다. 다시 말해, 플래너 모듈(205)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(205)은 복수의 동작, 및 복수의 컨셉 사이의 연관 정보(예: 온톨로지(ontology))가 포함된 플랜을 생성할 수 있다. 상기 플래너 모듈(205)은 컨셉과 동작의 관계들의 집합이 저장된 캡슐 데이터베이스(208)에 저장된 정보를 이용하여 플랜을 생성할 수 있다.
- [0051] 일 실시 예의 자연어 생성 모듈(206)은 지정된 정보를 텍스트 형태로 변경할 수 있다. 상기 텍스트 형태로 변경된 정보는 자연어 발화의 형태일 수 있다. 일 실시 예의 텍스트 음성 변환 모듈(207)은 텍스트 형태의 정보를 음성 형태의 정보로 변경할 수 있다.
- [0052] 일 실시 예에 따르면, 자연어 플랫폼(202)의 기능의 일부 기능 또는 전체 기능은 사용자 단말(220)에서도 구현 가능할 수 있다.
- [0053] 상기 캡슐 데이터베이스(208)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(208)는 CAN(concept action network)의 형태로 복수의 캡슐을 저장할 수 있다. 일 실시 예에 따르면, 복수의 캡슐은 캡슐 데이터베이스(208)에 포함된 기능 저장소(function registry)에 저장될 수 있다.
- [0054] 상기 캡슐 데이터베이스(208)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(208)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(208)는 사용자 단말(220)을 통해 출력되는 정보의 레이아웃(layout) 정보를 지

장하는 레이아웃 레지스트리(layout registry)를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(230)는 캡슐 정보에 포함된 어휘(vocabulary) 정보가 저장된 어휘 레지스트리(vocabulary registry)를 포함할 수 있다. 일 실시 예에 따르면, 캡슐 데이터베이스(230)는 사용자와의 대화(dialog)(또는, 인터랙션(interaction)) 정보가 저장된 대화 레지스트리(dialog registry)를 포함할 수 있다. 상기 캡슐 데이터베이스(208)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(208)은 사용자 단말(220) 내에도 구현이 가능할 수 있다.

[0055] 일 실시 예의 실행 엔진(209)은 상기 생성된 플랜을 이용하여 결과를 산출할 수 있다. 엔드 유저 인터페이스(210)는 산출된 결과를 사용자 단말(220)로 송신할 수 있다. 이에 따라, 사용자 단말(220)은 상기 결과를 수신하고, 상기 수신된 결과를 사용자에게 제공할 수 있다. 일 실시 예의 매니지먼트 플랫폼(211)은 지능형 서버(200)에서 이용되는 정보를 관리할 수 있다. 일 실시 예의 빅 데이터 플랫폼(212)은 사용자의 데이터를 수집할 수 있다. 일 실시 예의 분석 플랫폼(213)을 지능형 서버(200)의 QoS(quality of service)를 관리할 수 있다. 예를 들어, 분석 플랫폼(213)은 지능형 서버(200)의 구성 요소 및 처리 속도(또는, 효율성)를 관리할 수 있다.

[0056] 일 실시 예의 서비스 서버(230)는 사용자 단말(220)에 지정된 서비스(예: 음식 주문 또는 호텔 예약)를 제공할 수 있다. 일 실시 예에 따르면, 서비스 서버(230)는 제3 자에 의해 운영되는 서버일 수 있다. 일 실시 예의 서비스 서버(230)는 수신된 음성 입력에 대응되는 플랜을 생성하기 위한 정보를 지능형 서버(200)에 제공할 수 있다. 상기 제공된 정보는 캡슐 데이터베이스(208)에 저장될 수 있다. 또한, 서비스 서버(230)는 플랜에 따른 결과 정보를 지능형 서버(200)에 제공할 수 있다.

[0057] 위에 기술된 통합 지능 시스템에서, 상기 사용자 단말(220)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.

[0058] 일 실시 예에서, 상기 사용자 단말(220)은 내부에 저장된 지능형 앱(또는, 음성 인식 앱)을 통해 음성 인식 서비스를 제공할 수 있다. 이 경우, 예를 들어, 사용자 단말(220)은 상기 마이크를 통해 수신된 사용자 발화(utterance) 또는 음성 입력(voice input)을 인식하고, 인식된 음성 입력에 대응되는 서비스를 사용자에게 제공할 수 있다.

[0059] 일 실시 예에서, 사용자 단말(220)은 수신된 음성 입력에 기초하여, 단독으로 또는 상기 지능형 서버 및/또는 서비스 서버와 함께 지정된 동작을 수행할 수 있다. 예를 들어, 사용자 단말(220)은 수신된 음성 입력에 대응되는 앱을 실행시키고, 실행된 앱을 통해 지정된 동작을 수행할 수 있다.

[0060] 일 실시 예에서, 사용자 단말(220)이 지능형 서버(200) 및/또는 서비스 서버와 함께 서비스를 제공하는 경우에는, 상기 사용자 단말은, 상기 마이크(22)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(221)를 이용하여 지능형 서버(200)로 송신할 수 있다.

[0061] 일 실시 예에 따른 지능형 서버(200)는 사용자 단말(220)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.

[0062] 일 실시 예의 사용자 단말(220)은, 통신 인터페이스(221)를 이용하여 상기 응답을 수신할 수 있다. 사용자 단말(220)은 상기 스피커(223)를 이용하여 사용자 단말(220) 내부에서 생성된 음성 신호를 외부로 출력하거나, 디스플레이(224)를 이용하여 사용자 단말(220) 내부에서 생성된 이미지를 외부로 출력할 수 있다.

[0063] 도 2B는 다양한 실시 예에 따른, 컨셉과 동작의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.

- [0064] 상기 지능형 서버(200)의 캡슐 데이터베이스(예: 캡슐 데이터베이스(208))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.
- [0065] 상기 캡슐 데이터베이스는 복수의 도메인(예: 어플리케이션) 각각에 대응되는 복수의 캡슐(capsule(A)(241), capsule(B)(244))을 저장할 수 있다. 일 실시 예에 따르면, 하나의 캡슐(예: capsule(A)(241))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(242) 또는 CP 2 (243))가 대응될 수 있다. 일 실시 예에 따르면, 하나의 캡슐은 지정된 기능을 수행하기 위한 적어도 하나 이상의 동작(248) 및 적어도 하나 이상의 컨셉(249)을 포함할 수 있다.
- [0066] 상기, 자연어 플랫폼(220)은 캡슐 데이터베이스에 저장된 캡슐을 이용하여 수신된 음성 입력에 대응하는 태스크를 수행하기 위한 플랜을 생성할 수 있다. 예를 들어, 자연어 플랫폼의 플래너 모듈(202)은 캡슐 데이터베이스에 저장된 캡슐을 이용하여 플랜을 생성할 수 있다. 예를 들어, 캡슐 A (241)의 동작들(241_1, 241_3)과 컨셉들(241_2, 241_4) 및 캡슐 B(244)의 동작(244_1)과 컨셉(244_2)를 이용하여 플랜(247)을 생성할 수 있다.
- [0067] 도 2C는 다양한 실시 예에 따른 사용자 단말이 지능형 앱을 통해 수신된 음성 입력을 처리하는 화면을 나타낸 도면이다.
- [0068] 사용자 단말(220)은 지능형 서버(200)를 통해 사용자 입력을 처리하기 위해 지능형 앱을 실행할 수 있다.
- [0069] 일 실시 예에 따르면, 250 화면에서, 사용자 단말(220)은 지정된 음성 입력(예: 웨이크 업!)을 인식하거나 하드웨어 키(예: 전용 하드웨어 키)를 통한 입력을 수신하면, 음성 입력을 처리하기 위한 지능형 앱을 실행할 수 있다. 사용자 단말(220)은, 예를 들어, 스케줄 앱을 실행한 상태에서 지능형 앱을 실행할 수 있다. 일 실시 예에 따르면, 사용자 단말(220)은 지능형 앱에 대응되는 오브젝트(예: 아이콘)(251)를 디스플레이(224)에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(220)은 사용자 발화에 의한 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(220)은 "이번 주 일정 알려줘!"라는 음성 입력을 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말(220)은 수신된 음성 입력의 텍스트 데이터가 표시된 지능형 앱의 UI(user interface)(253)(예: 입력 창)를 디스플레이에 표시할 수 있다.
- [0070] 일 실시 예에 따르면, 260 화면에서, 사용자 단말(220)은 수신된 음성 입력에 대응되는 결과를 디스플레이에 표시할 수 있다. 예를 들어, 사용자 단말(220)은 수신된 사용자 입력에 대응되는 플랜을 수신하고, 플랜에 따라 '이번 주 일정'을 디스플레이에 표시할 수 있다.
- [0071] 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 한정되지 않는다.
- [0072] 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. 아이템에 대응하는 명사의 단수 형은 관련된 문맥상 명백하게 다르게 지시하지 않는 한, 상기 아이템 한 개 또는 복수개를 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.
- [0073] 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, 일 실시예에 따르면, 모듈은 ASIC(application-specific integrated circuit)의 형태로 구현될 수 있다.

- [0074] 본 문서의 다양한 실시예들은 기기(machine)(예: 전자 장치(101)) 의해 읽을 수 있는 저장 매체(storage medium)(예: 내장 메모리(136) 또는 외장 메모리(138))에 저장된 하나 이상의 명령어들을 포함하는 소프트웨어(예: 프로그램(140))로서 구현될 수 있다. 예를 들면, 기기(예: 전자 장치(101))의 프로세서(예: 프로세서(120))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: 전자 기파)를 포함하지 않는다는 것을 의미할 뿐이며, 이 용어는 데이터가 저장매체에 반영구적으로 저장되는 경우와 임시적으로 저장되는 경우를 구분하지 않는다.
- [0075] 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.
- [0076] 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.
- [0077] 도 3은 본 발명의 다양한 실시 예에 따른 지능형 에이전트 시스템을 개략적으로 나타낸 도면이다.
- [0078] 다양한 실시 예에 따르면, 전자 장치(101)는 지능형 에이전트 시스템(300)을 이용하여, 사용자의 명령에 따른 기능을 수행할 수 있다. 예를 들어, 지능형 에이전트 시스템(300)은 상황 정보 확인 모듈(310), 음성 정보 인식 모듈(320), 영상 정보 인식 모듈(330), 기능 수행 모듈(340) 등을 포함할 수 있다. 도 3의 지능형 에이전트 시스템(300)은 도 2A 내지 2C에 개시된 통합 지능 시스템의 적어도 일부 구성 또는 모듈을 포함하거나, 적어도 일부 구성 또는 모듈의 조합에 해당할 수 있다.
- [0079] 다양한 실시 예에 따르면, 상황 정보 확인 모듈(310)은 전자 장치(101)와 관련된 적어도 하나의 상황 정보를 확인할 수 있다. 예를 들어, 상황 정보는 전자 장치(101) 주변의 소음 정도를 포함할 수 있다. 상황 정보 확인 모듈(310)은 전자 장치(101)의 사용자가 콘서트에 참석한 경우, 전자 장치(101) 주변이 매우 시끄러워 소음이 심하다는 것을 상황 정보로 획득 및 저장할 수 있다. 다른 예를 들어, 상황 정보는 전자 장치(101)에서 실행 중인 어플리케이션에 관한 정보를 포함할 수 있다. 상황 정보 확인 모듈(310)은 전자 장치(101)에서 동영상 또는 음악을 재생 중인 경우와 같이 실행 중인 어플리케이션에 관한 정보를 상황 정보로 획득할 수 있다. 상황 정보 확인 모듈(310)은 특정 어플리케이션과 관련하여, 특정 어플리케이션의 동작 상태에 관한 정보를 상황 정보로 획득할 수 있다. 예를 들어, 상황 정보 확인 모듈(310)은 특정 어플리케이션이 동영상 어플리케이션이라면, 현재 동영상이 재생 중인지, 또는 멈춤 상태인지 등을 확인할 수 있다. 다른 예를 들어, 상황 정보 확인 모듈(310)은 특정 어플리케이션이 포어그라운드(foreground) 또는 백그라운드(background)에서 동작 중인지 확인할 수 있다. 전자 장치(101)는 상황 정보 확인 모듈(310)에서 획득한 적어도 하나의 상황 정보는 전자 장치(101)에서 지능형 에이전트 기능을 제공하는데 활용될 수 있다.
- [0080] 다양한 실시 예에 따르면, 음성 정보 인식 모듈(320)은 사용자의 음성을 수신할 수 있으며, 수신된 음성을 기반으로 사용자의 명령이 무엇인지 판단할 수 있다. 예를 들어, 음성 정보 인식 모듈(320)은 자동 발화 인식(automatic speech recognition, ASR)을 이용하여 음성을 텍스트로 변환하고, 변환된 텍스트에 기반하여 기능을 수행할 수 있다.

- [0081] 다양한 실시 예에 따르면, 전자 장치(101)의 주변 소음이 심한 경우에는 사용자로부터 음성을 원활하게 수신할 수 없다는 문제가 발생할 수 있다. 다른 예를 들어, 사용자가 위치하는 곳이 도서관과 같이 정숙을 요하는 곳인 경우, 사용자가 큰 소리로 음성을 입력 하는 것이 곤란한 경우가 발생할 수 있다. 상기와 같은 경우에, 전자 장치(101)에서 음성 입력을 통한 지능형 에이전트 서비스를 수행하는 것이 곤란할 수 있다.
- [0082] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 전자 장치(101)의 카메라 모듈(180)을 통해 사용자의 영상 정보를 수신할 수 있으며, 수신된 영상 정보를 기반으로 립 리딩(lip reading)을 수행할 수 있다. 예를 들어, 립 리딩은 화자의 입술 움직임을 분석하여 단어 또는 문장을 인식하는 기술을 의미할 수 있다. 영상 정보 인식 모듈(330)은 훈련된 모델을 통해, 수신된 립 영상을 기반으로 가장 적합한 단어 또는 문장을 찾는 방법을 이용할 수 있다. 이에 따라, 음성 정보 인식 모듈(320)과는 달리 텍스트를 생성 하지 않을 수도 있다.
- [0083] 상기 서술한 것과 같이 사용자의 음성만으로 사용자가 수행하고자 하는 기능을 확인하기가 어려운 경우에, 수신된 영상 정보를 기반으로 립 리딩을 수행할 수 있으며, 이를 통해 지능형 에이전트 서비스의 정확도를 향상 시킬 수 있다.
- [0084] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 상황 정보 확인 모듈(310)에서 획득한 상황 정보에 기반하여, 음성 정보의 수신 없이, 사용자의 립 리딩만으로 사용자 명령을 확인하고, 확인된 사용자 명령에 상응하는 기능을 수행할 수 있다.
- [0085] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 립 리딩을 통해 음성 정보 인식 모듈(320)에서 처리될 음성 정보를 보완할 수 있다. 전자 장치(101)는 음성 정보 및 립 리딩 정보를 조합하여 사용자의 명령을 확인할 수 있으며, 확인된 사용자 명령에 상응하는 기능을 수행할 수 있다.
- [0086] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 사용자의 입술 모양에 기반하여, 영상에 포함된 입술의 개수, 음성 명령을 입력 중인 사용자, 발화의 시작점(start point) 또는 종료점(end point) 등에 관한 정보를 확인할 수 있다.
- [0087] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 립 리딩과 관련된 적어도 하나의 정보를 사용자 인터페이스로 제공할 수 있다. 예를 들어, 영상 정보 인식 모듈(330)은 립 리딩을 진행 중인지 여부에 관한 정보, 영상에 포함된 입술에 관한 정보, 립 리딩 중인 입술에 관한 정보, 또는 립 리딩 오류에 관한 정보 등을 사용자 인터페이스로 제공할 수 있다.
- [0088] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 립 리딩을 원활하게 하기 위해 립 리딩에 관한 가이드를 제공할 수 있다. 예를 들어, 영상 정보 인식 모듈(330)은 사용자의 입술 부분을 클로즈 업 해서 별도의 프리뷰 화면으로 제공할 수 있다. 다른 예를 들어, 수신된 영상에서 사용자의 입술을 파악하기 힘든 경우, 전자 장치(101)의 배향 등을 조절하도록 안내할 수 있다.
- [0089] 다양한 실시 예에 따르면, 영상 정보 인식 모듈(330)은 영상에 포함된 입술이 복수인 경우, 현재 발화 중이라고 인식하는 입술과 단지 영상에 포함된 입술을 구분하여 표시할 수 있다. 예를 들어, 실제로 음성 명령을 입력하고 있다고 판단되는 입술의 경우, 입술 주변 영역을 테두리로 표시할 수 있으며, 단지 영상에 포함된 입술에 대해서는 별도의 표시를 하지 않을 수 있다. 다른 예를 들어, 실제로 음성 명령을 입력하고 있다고 판단되는 입술의 경우, 입술 주변 영역을 제1 테두리로 표시할 수 있으며, 단지 영상에 포함된 입술에 대해서는 제2 테두리로 표시할 수 있다. 여기서, 제1 테두리는 제2 테두리와 선의 형태(예: 실선, 점선 등), 색채, 굵기 등이 상이할 수 있다.
- [0090] 다양한 실시 예에 따르면, 기능 수행 모듈(340)은 전자 장치(101)의 상황 정보, 음성 정보 또는 영상 정보 중 적어도 하나에 기반하여 사용자의 명령이 무엇인지 결정할 수 있으며, 결정된 명령에 상응하는 기능을 수행할 수 있다.
- [0091] 도 4는 본 발명의 다양한 실시 예에 따른 립 리딩을 수행하기 위한 전자 장치의 블록도이다.
- [0092] 다양한 실시 예에 따르면, 전자 장치(400, 예: 도 1의 전자 장치 101, 도 2A의 사용자 단말 220)는 립 리딩을 수행하기 위한 복수의 모듈로 구분될 수 있다.
- [0093] 다양한 실시 예에 따르면, 클라이언트(client, 410) 모듈은 지능형 에이전트 서비스에 접속하고, 지능형 에이전트 서비스 결과를 출력하는 역할을 수행할 수 있다. 예를 들어, 클라이언트(410)는 현재 수행 중인 어플리케이션에 관한 정보 및 전자 장치(400)의 주변 상황에 따라 립 리딩 기능을 실행할 수 있다.

- [0094] 다양한 실시 예에 따르면, 자동 음성 인식(automatic speech recognition, ASR, 420) 모듈은 입력된 음성 신호를 텍스트로 변환할 수 있다.
- [0095] 다양한 실시 예에 따르면, 입술 확인(lip verification, 431) 모듈은 카메라 모듈(180)을 통해 획득한 영상 정보에서 입술이 감지되는지 여부를 확인하고, 입술의 움직임이 인식 가능한 상태인지 확인할 수 있다. 예를 들어, 입술 확인 모듈(431)은 카메라 모듈을 통해 획득하는 영상 정보를 기반으로 입술 인식을 할 수 있는 영상인지 여부를 판단하여 입술 인식을 계속할지, 또는 음성 인식만을 수행할지를 결정할 수 있다.
- [0096] 다양한 실시 예에 따르면, 입술 인식(lip recognizer, 433) 모듈은 인식된 입술 영역에 대해, 일반 립 리딩 모델(general lip reading model, 435)과 개인 립 리딩 모델(personal lip reading model, 437)을 기반으로 입술의 움직임을 감지하여 사용자 발화를 인식 또는 추정할 수 있다.
- [0097] 다양한 실시 예에 따르면, 일반 립 리딩 모델(435)은 불특정 다수의 입술 움직임과 이에 상응하는 발화 정보를 기반으로 학습한 모델로서 일반적인 립 리딩의 인식 성능을 확보할 수 있다.
- [0098] 다양한 실시 예에 따르면, 개인 립 리딩 모델(437)은 사용자의 입술 움직임을 학습하여 모델을 생성함으로써 전자 장치(400)의 사용자에게 최적화된 인식 성능을 확보할 수 있다. 예를 들어, 사람마다 발화 습관, 입술 움직임이 상이하기 때문에 개인별 성능 열화가 발생할 수 있는데, 개인 립 리딩 모델(437)을 통해 립 리딩의 인식 성능을 향상시킬 수 있다.
- [0099] 다양한 실시 예에 따르면, 전자 장치(400)를 통해, 하기와 같은 동작들을 수행할 수 있다. 예를 들어, 지능형 에이전트 서비스가 사용자의 입력(예: 물리적 키 입력, 웨이크 업(wake up) 발화 등)에 의해 실행되는 경우, 실행 중인 어플리케이션(441, 443)의 상황 정보를 수신한다. 예를 들어, 동영상 또는 음악이 재생되고 있는 상황이면 입술 인식을 위해 카메라 모듈을 활성화할 수 있다. 카메라 모듈로부터 수집된 영상 정보를 기반으로, 입술 인식을 할 수 있는 상태임이 결정되면 수신되는 사용자 발화 및 입술 움직임을 통해 사용자의 발화가 무엇인지 결정할 수 있다.
- [0100] 도 5는 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 사용자 명령에 상응하는 기능을 수행하는 방법(500)에 관한 흐름도이다.
- [0101] 다양한 실시 예에 따르면, 전자 장치(101)는 510 동작에서, 지능형 에이전트 서비스를 실행할 수 있다. 예를 들어, 전자 장치(101)는 사용자의 입력(예: 물리적 키 입력, 웨이크 업(wake up) 발화 등)을 수신하여, 지능형 에이전트 서비스를 실행할 수 있다.
- [0102] 다양한 실시 예에 따르면, 전자 장치(101)는 520 동작에서, 전자 장치(101)의 상황 정보를 확인할 수 있다. 예를 들어, 상황 정보는 전자 장치(101) 주변의 소음 정도를 포함할 수 있다. 다른 예를 들어, 상황 정보는 전자 장치(101)에서 실행 중인 어플리케이션에 관한 정보를 포함할 수 있다.
- [0103] 다양한 실시 예에 따르면, 전자 장치(101)는 530 동작에서, 확인된 상황 정보에 기반하여, 영상 정보를 획득할 것인지 결정할 수 있다. 예를 들어, 전자 장치(101)의 주변 소음이 심한 경우에는 사용자로부터 음성을 원활하게 수신할 수 없기에 영상 정보를 획득하도록 제어할 수 있다. 다른 예를 들어, 사용자가 위치하는 곳이 정숙을 요하는 곳인 경우, 사용자가 큰 소리로 음성을 입력 하는 것이 곤란할 수 있다. 이와 같은 경우에, 영상 정보를 추가로 획득하여 지능형 에이전트 서비스의 정확도를 향상시킬 수 있다.
- [0104] 다양한 실시 예에 따르면, 전자 장치(101)는 영상 정보를 획득하는 것으로 결정한 경우, 541 동작으로 분기하여, 카메라 모듈을 활성화하여 사용자의 영상 정보를 획득할 수 있다.
- [0105] 다양한 실시 예에 따르면, 전자 장치(101)는 543 동작에서, 획득한 영상 정보에 기반하여, 사용자 발화를 인식할 수 있다.
- [0106] 다양한 실시 예에 따르면, 전자 장치(101)는 영상 정보를 획득하지 않는 것으로 결정한 경우, 551 동작으로 분기하여, 사용자로부터 음성 정보를 획득할 수 있다.
- [0107] 다양한 실시 예에 따르면, 전자 장치(101)는 553 동작에서, 음성 정보에 기반하여, 사용자 발화를 인식할 수 있다.
- [0108] 다양한 실시 예에 따르면, 전자 장치(101)는 560 동작에서, 생성된 텍스트에 상응하는 기능을 수행할 수 있다.
- [0109] 다양한 실시 예에 따르면, 도 5를 참조하면, 전자 장치(101)는 영상 정보 또는 음성 정보를 선택적으로 획득하

여, 사용자 발화에 상응하는 기능을 실행할 수 있다. 다만, 본 개시는 이러한 실시 예에 그치지 않으며, 영상 정보 및 음성 정보를 함께 획득하고, 이들의 조합에 기반하여 사용자 발화(예: 조합 발화)에 상응하는 기능을 실행하는 실시 예도 포함할 수 있다.

- [0110] 도 6은 본 발명의 다양한 실시 예에 따른 전자 장치에서 사용자의 음성 또는 사용자의 입술 모양 중 적어도 하나에 기반하여, 사용자 발화를 인식하는 방법을 나타내는 도면이다.
- [0111] 다양한 실시 예에 따르면, 전자 장치(101)는 610 동작에서, 상황 정보에 기반하여, 립 리딩을 실행할지 여부를 결정할 수 있다. 만약, 립 리딩을 실행하기로 결정한 경우, 621 동작으로 분기하여 사용자의 영상 정보를 획득할 수 있다. 만약 립 리딩을 실행하지 않기로 결정한 경우, 623 동작으로 분기하여 음성 정보를 획득할 수 있다. 다른 예를 들어, 전자 장치(101)는 립 리딩을 실행하기로 결정한 경우에도, 623 동작으로 분기하여 음성 정보를 추가로 획득할 수 있다.
- [0112] 다양한 실시 예에 따르면, 전자 장치(101)는 립 리딩을 실행하여 영상 정보를 획득한 경우, 입술에 관한 정보(631)를 확인할 수 있다. 예를 들어, 입술에 관한 정보(631)를 확인하는 방법은 입술 감지(lip detection), 입술 활성화 감지(lip activation detection), 입술 움직임 추적(lip tracking) 등을 포함할 수 있다. 예를 들어, 입술 감지는 영상 정보에서 사용자의 입술이 무엇인지 감지하는 것을 의미할 수 있다. 예를 들어, 입술 활성화 감지는 입술이 움직이기 시작함을 감지하는 것을 의미할 수 있다. 예를 들어, 입술 움직임 추적은 움직이기 시작한 입술이 어떠한 모양으로 변화하는지 지속적으로 추적하는 것을 의미할 수 있다.
- [0113] 다양한 실시 예에 따르면, 전자 장치(101)는 입술에 관한 정보(631)에 기반하여 사용자의 발화를 인식할 수 있다. (예: 입술 움직임만으로 사용자 발화를 인식)
- [0114] 다양한 실시 예에 따르면, 전자 장치(101)는 자동 음성 인식 모듈(633)을 통해 사용자의 음성을 텍스트로 변환(645)할 수 있다. (예: 수신된 음성만으로 사용자 발화를 텍스트로 변환)
- [0115] 다양한 실시 예에 따르면, 전자 장치(101)는 입술에 관한 정보(631) 및 자동 음성 인식 모듈(633)로부터 획득한 정보를 조합하여 사용자의 발화를 인식(641)할 수 있다. (예: 수신된 음성 및 입술 움직임을 조합하여 사용자 발화를 인식)
- [0116] 다양한 실시 예에 따르면, 전자 장치(101)에서 립 리딩을 실행하여 영상 정보를 획득하는 경우, 사용자의 발화를 인식하기 위해, 데이터 베이스(643)를 참조할 수 있다. 해당 데이터 베이스(643)는 서버 또는 전자 장치(101) 내부 메모리(130)에 저장되어 있을 수 있다.
- [0117] 다양한 실시 예에 따르면, 전자 장치(101)는 650 동작에서, 변환된 텍스트에 상응하는, 사용자의 발화를 인식할 수 있다..
- [0118] 도 7은 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 기능을 수행하는 경우의 사용자 인터페이스를 나타내는 도면이다.
- [0119] 다양한 실시 예에 따르면, 도 710을 참조하면, 전자 장치(101)는 제1 어플리케이션(예: 음악 재생 어플리케이션)을 실행할 수 있다. 전자 장치(101)는 제1 어플리케이션을 실행하는 중, 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신할 수 있다.
- [0120] 다양한 실시 예에 따르면, 도 720을 참조하면, 전자 장치(101)는 제1 어플리케이션(예: 음악 재생 어플리케이션)이 지능형 에이전트 서비스가 활성화된 상태에서 끊임 없이 계속해서 백그라운드로 실행 중인 경우(예: 음악 재생 중), 지능형 에이전트 서비스를 통해 사용자의 음성 또는 입술에 관한 정보 중 적어도 하나를 수신할 수 있다. 예를 들어, 전자 장치(101)는 지능형 에이전트 서비스가 활성화 되어 음성 또는 입술에 관한 정보 중 적어도 하나를 수신하고 있다는 것을 별도의 객체 또는 아이콘(721)으로 표시할 수 있다.
- [0121] 다양한 실시 예에 따르면, 도 720을 참조하면, 전자 장치(101)는 립 리딩을 하기 위해 카메라 모듈(180)을 활성화한 경우 이를 사용자에게 알리기 위한 별도의 객체 또는 아이콘(723)을 표시할 수 있다. 예를 들어, 립 리딩 객체 또는 아이콘(723)은 립 리딩의 상태에 따라 상태가 변화되어 표시될 수 있다. 예를 들어, 사용자의 입술 움직임을 정확하게 읽고 있는 경우에는, 객체 또는 아이콘(723)을 선명하게 표시할 수 있다. 만약, 사용자의 입술 움직임을 파악하기 어려운 경우(예: 입술 위치 감지 실패, 주변 밝기 어두움 등)에는 객체 또는 아이콘(723)을 흐릿하게 표시하거나 깜빡이게 할 수 있다. 다른 예를 들어, 사용자의 입술 움직임을 정확하게 읽고 있는 경우에는, 객체 또는 아이콘(723)을 제1 색상으로 표시할 수 있다. 만약, 사용자의 입술 움직임을 파악하기 어려운 경우(예: 입술 위치 감지 실패, 주변 밝기 어두움 등)에는 객체 또는 아이콘(723)을 제1 색상과 다른 제2

색상으로 표시할 수 있다.

- [0122] 다양한 실시 예에 따르면, 도 720을 참조하면, 전자 장치(101)는 사용자의 발화를 텍스트로 변환한 결과를 실시간으로 표시할 수 있다. 예를 들어, 전자 장치(101)는 사용자가 발화하는 중에도 중간 결과(725)를 사용자에게 제공할 수 있으며, 사용자가 발화를 마친 경우 최종 결과(727)를 사용자에게 제공할 수 있다.
- [0123] 다양한 실시 예에 따르면, 도 720을 참조하면, 전자 장치(101)는 사용자로부터 명령을 수신하기 위해, 현재 음악 재생 어플리케이션을 통해 재생 중인 음악을 잠시 멈출 수 있다. 이 경우, 재생 중인 음악이 멈추었기 때문에 사용자의 음성 입력을 보다 원활하게 수신할 수 있다. 다만, 사용자의 음악 감상이 연속적이지 않기에, 사용자의 불편함을 초래할 여지가 있다.
- [0124] 다양한 실시 예에 따르면, 도 730을 참조하면, 전자 장치(101)는 사용자의 발화가 제2 어플리케이션(예: 갤러리)을 실행하라는 것인 경우, 제2 어플리케이션으로 화면을 전환할 수 있다. 이후, 전자 장치(101)는 사용자로부터 뒤로 가기 입력을 수신한 경우에는 재차 제1 어플리케이션의 실행 화면으로 복귀할 수 있다.
- [0125] 도 8A 내지 8B는 본 발명의 다양한 실시 예에 따른 전자 장치의 지능형 에이전트 서비스를 통해 기능을 수행하는 경우의 사용자 인터페이스를 나타내는 도면이다.
- [0126] 다양한 실시 예에 따르면, 도 8A를 참조하면, 전자 장치(101)는 810 화면에서, 동영상 재생 어플리케이션을 통해 영화를 재생할 수 있다. 이러한 상황에서, 전자 장치(101)는 사용자로부터 지능형 에이전트 서비스를 활성화하는 입력을 수신할 수 있다. 전자 장치(101)는 재생 중인 영화를 멈춤(pause 또는 stop)하지 않은 채 사용자로부터 입력을 수신할 수 있다. 예를 들어, 전자 장치(101)는 현재 실행 중인 어플리케이션이 동영상 재생 어플리케이션이라는 것과, 현재 동영상 재생 어플리케이션을 통해 영화가 재생 중이라는 상황 정보를 확인할 수 있다. 전자 장치(101)는 사용자의 영화 감상을 방해하지 않기 위해, 동영상 재생을 중단하지 않은 채 립 리딩 모드를 활성화하여, 사용자의 입술 움직임을 감지할 수 있다. 물론, 이 경우에도, 사용자의 음성을 함께 수신할 수 있으며, 전자 장치(101)는 사용자의 음성 및 입술 움직임을 기반으로 사용자가 실행하려고 하는 명령을 인식할 수 있다. 한편, 전자 장치(101)는 립 리딩 모드가 활성화 되었음을 알리는 객체 또는 아이콘(811)을 사용자 인터페이스로 표시할 수 있다.
- [0127] 다양한 실시 예에 따르면, 도 8A를 참조하면, 전자 장치(101)는 820 화면에서, 사용자로부터 금요일 날씨가 어떻게 되는지를 묻는 입력을 수신할 수 있다. 이에 대응하여, 전자 장치(101)는 "금요일에는 눈송이가 날릴 예정입니다."라는 대답을 제공할 수 있다. 사용자가 현재 영화를 감상 중이기 때문에, 전자 장치(101)가 별도의 음성을 통해 대답을 하거나, 현재 실행 중인 어플리케이션을 가리고 지능형 에이전트 서비스 화면으로 이동을 하게 되면 원활한 영화 감상을 방해할 수 있다. 이에 따라, 전자 장치(101)는 별도의 음성 대답이나 화면 이동 없이, 자막과 유사한 형태로 사용자의 명령에 상응하는 대답(821)을 표시할 수 있다.
- [0128] 다양한 실시 예에 따르면, 도 8B를 참조하면, 전자 장치(101)는 830 화면에서, 음악 재생 어플리케이션을 통해 음악을 재생할 수 있다. 이러한 상황에서, 전자 장치(101)는 사용자로부터 지능형 에이전트 서비스를 활성화하는 입력을 수신할 수 있다. 전자 장치(101)는 재생 중인 음악을 멈춤(pause 또는 stop)하지 않은 채 사용자로부터 입력을 수신할 수 있다. 예를 들어, 전자 장치(101)는 현재 실행 중인 어플리케이션이 음악 재생 어플리케이션이라는 것과, 현재 음악 재생 어플리케이션을 통해 음악이 재생 중이라는 상황 정보를 확인할 수 있다. 전자 장치(101)는 사용자의 음악 감상을 방해하지 않기 위해, 음악 재생을 중단하지 않은 채 립 리딩 모드를 활성화하여, 사용자의 입술 움직임을 감지할 수 있다. 물론, 이 경우에도, 사용자의 음성을 함께 수신할 수 있으며, 전자 장치(101)는 사용자의 음성 및 입술 움직임을 기반으로 사용자가 실행하려고 하는 명령을 인식할 수 있다.
- [0129] 다양한 실시 예에 따르면, 도 8B를 참조하면, 전자 장치(101)는 840 화면에서, 사용자로부터 메시지를 보내달라는 요청을 인식할 수 있다. 이에 대응하여, 전자 장치(101)는 누구에게 보낼까요? 라는 의미의 "To whom(847)"이라는 대답(841)을 제공할 수 있다. 이와 더불어, 전자 장치(101)는 현재 사용자의 피드백을 수신 중이라는 의미의 인디케이터(indicator, 843)를 표시하고, 립 리딩을 실행하고 있다는 의미의 객체 또는 아이콘(845)을 표시할 수 있다. 여기서, 인디케이터, 객체, 또는 아이콘은 무엇인가를 지칭하기 위해 사용된 그래픽 유저 인터페이스로서 서로 유사한 의미로 사용된 것이며, 얼마든지 다른 표현으로 대체 가능할 수 있다. 도 8A와 마찬가지로, 사용자가 현재 음악을 감상 중이기 때문에, 전자 장치(101)가 별도의 음성을 통해 대답을 하거나 현재 실행 중인 어플리케이션을 가리고 지능형 에이전트 서비스 화면으로 이동을 하게 되면 원활한 영화 감상을 방해할 수 있다. 이에 따라, 전자 장치(101)는 별도의 음성 대답 없이, 메시지와 유사한 형태로 현재 실행 중인 어플리케이션이 디스플레이 되는 일부 영역에 사용자의 명령에 상응하는 대답(841)을 표시할 수 있다.

- [0130] 도 9는 본 발명의 다양한 실시 예에 따른 전자 장치의 립 리딩이 가능한지 여부에 기반하여 사용자 명령에 상응하는 기능을 수행하는 방법(900)에 관한 흐름도이다.
- [0131] 다양한 실시 예에 따르면, 전자 장치(101)는 910 동작에서, 사용자의 영상 정보를 획득할 수 있다.
- [0132] 다양한 실시 예에 따르면, 전자 장치(101)는 920 동작에서, 획득한 영상 정보에서 입술을 인식할 수 있는지(예: 립 리딩이 가능한지) 판단할 수 있다.
- [0133] 다양한 실시 예에 따르면, 전자 장치(101)는 입술을 인식할 수 있는 경우, 930 동작으로 분기하여, 립 리딩에 관한 사용자 인터페이스(user interface, UI)를 표시할 수 있다.
- [0134] 다양한 실시 예에 따르면, 전자 장치(101)는 940 동작에서, 립 리딩을 수행하고, 립 리딩에 관한 중간 결과를 출력할 수 있다.
- [0135] 다양한 실시 예에 따르면, 전자 장치(101)는 950 동작에서, 사용자의 명령 입력이 완료되었는지 판단할 수 있다.
- [0136] 다양한 실시 예에 따르면, 전자 장치(101)는 사용자의 명령 입력이 완료된 경우, 960 동작으로 분기하여, 립 리딩의 최종 결과를 출력할 수 있다.
- [0137] 다양한 실시 예에 따르면, 전자 장치(101)는 사용자의 명령 입력이 완료되지 않은 경우, 940 동작으로 분기하여, 계속해서 립 리딩을 수행하고, 립 리딩에 관한 중간 결과를 출력할 수 있다.
- [0138] 다양한 실시 예에 따르면, 전자 장치(101)는 입술을 인식할 수 없는 경우, 935 동작으로 분기하여, 음성 인식에 관한 사용자 인터페이스(UI)를 표시할 수 있다.
- [0139] 다양한 실시 예에 따르면, 전자 장치(101)는 945 동작에서, 음성 인식을 수행하고, 음성 인식에 관한 중간 결과를 출력할 수 있다.
- [0140] 다양한 실시 예에 따르면, 전자 장치(101)는 955 동작에서, 사용자의 명령 입력이 완료되었는지 판단할 수 있다.
- [0141] 다양한 실시 예에 따르면, 전자 장치(101)는 사용자의 명령 입력이 완료된 경우, 965 동작으로 분기하여, 음성 인식의 최종 결과를 출력할 수 있다.
- [0142] 다양한 실시 예에 따르면, 전자 장치(101)는 사용자의 명령 입력이 완료되지 않은 경우, 945 동작으로 분기하여, 계속해서 음성 인식을 수행하고, 음성 인식에 관한 중간 결과를 출력할 수 있다.
- [0143] 다양한 실시 예에 따르면, 전자 장치(101)는 970 동작에서, 립 리딩 최종 결과 또는 음성 인식 최종 결과 중 적어도 하나에 기반하여 입력 명령을 수행할 수 있다.
- [0144] 도 10A 내지 10C는 본 발명의 다양한 실시 예에 따른 전자 장치의 립 리딩 동작에 관한 사용자 인터페이스를 제공하는 방법을 개략적으로 나타낸 도면이다.
- [0145] 다양한 실시 예에 따르면, 도 10A를 참조하면, 전자 장치(101)는 주변 소음이 일정 수준 이상으로 높거나 여러 사람들의 목소리가 동시에 입력될 경우 립 리딩을 실행할 수 있다. 전자 장치(101)는 입술 인식(lip recognizer) 모듈 및 자동 음성 인식(automatic speech recognition) 모듈을 통해 수신된 사용자의 발화가 무엇인지 인식할 수 있다. 예를 들어, 전자 장치(101)는 영상 정보에 포함된 입술의 움직임(도 1010 내지 1060)을 감지할 수 있으며, 이를 기반으로 사용자의 발화를 인식할 수 있다.
- [0146] 다양한 실시 예에 따르면, 도 10B를 참조하면, 전자 장치(101)는 립 리딩을 수행하기 위한 영상 정보가 기 설정된 밝기 이하의 것인 경우, 립 리딩의 정확도가 떨어지기에 사용자에게 좀 더 밝은 곳으로의 이동을 요청할 수 있다. 예를 들어, 어두운 곳에서 사용자에게 관한 이미지(1070)를 획득하는 경우, 사용자 입술 주변에 마크(1071)를 표시할 수 있다. 예를 들어, 마크(1071)는 정상적으로 립 리딩을 수행 중인 경우에는 실선으로 나타낼 수 있으며, 립 리딩을 수행하기 어려운 조건인 경우에는 점선으로 나타낼 수 있다. 다른 예를 들어, 마크(1071)는 정상적으로 립 리딩을 수행 중인 경우에는 녹색으로 나타낼 수 있으며, 립 리딩을 수행하기 어려운 조건인 경우에는 붉은색으로 나타낼 수 있다.
- [0147] 다양한 실시 예에 따르면, 도 10C를 참조하면, 전자 장치(101)는 영상 정보(1080)로부터 복수 개의 입술들을 감지할 수 있다. 전자 장치(101)는 감지된 입술 중에서 어떠한 입술이 현재 음성 명령을 입력하고 있는 입술인지

판단할 수 있다. 예를 들어, 전자 장치(101)는 사용자의 입술 모양의 움직임과 입력 중인 음성을 비교할 수 있으며, 이를 기반으로 복수 개의 입술들 중 실제로 음성 명령을 입력 중인 입술을 판단할 수 있다.

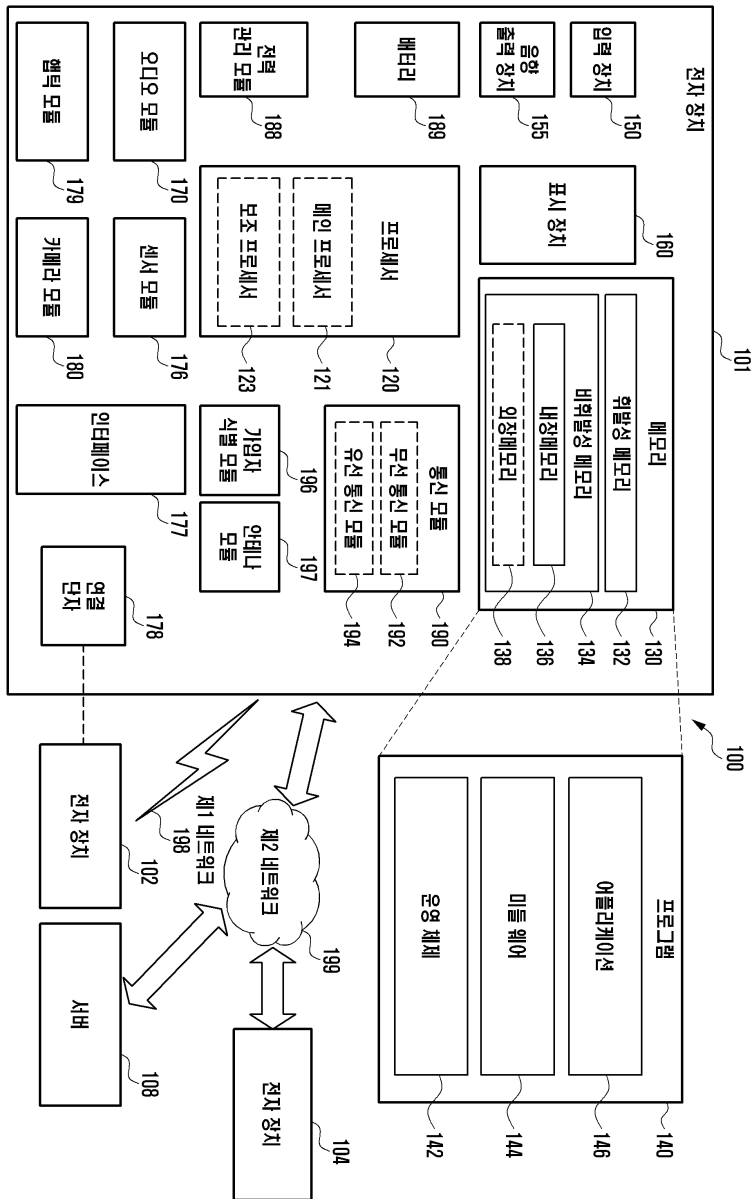
- [0148] 다양한 실시 예에 따르면, 도 10C를 참조하면, 전자 장치(101)는 영상 정보(1080)에 현재 음성 명령을 입력 중인 입술(1081)을 다른 입술들(1082)과 구분하여 표시할 수 있다. 예를 들어, 현재 음성 명령을 입력 중인 입술(1081)은 굵은 실선으로 표시하고, 다른 입술들(1082)은 연한 점선으로 표시할 수 있다.
- [0149] 다양한 실시 예에 따르면, 전자 장치(101)는 수신된 영상 정보의 상태에 따라, 음성 인식 및 립 리딩의 가중치를 조절할 수 있다. 예를 들어, 수신된 영상의 밝기가 기 설정된 밝기 이하인 경우, 립 리딩의 정확도가 낮다고 판단할 수 있다. 이에 따라, 음성 인식 및 립 리딩을 조합하여 사용자의 명령을 분석하는 경우, 음성 인식을 통해 획득한 결과에 가중치를 높게 할당할 수 있으며, 립 리딩을 통해 획득한 결과는 상대적으로 낮은 가중치를 할당할 수 있다.
- [0150] 다양한 실시 예에 따르면, 전자 장치(101)는 수신된 음성 정보가 실제로 음성 명령을 입력 중인 사용자의 것이 아닌 경우, 음성 인식 및 립 리딩의 가중치를 조절할 수 있다. 예를 들어, 실제 음성 명령을 입력 중인 사용자의 음성과 더불어 다른 사람들의 음성을 함께 수신 중인 경우, 음성 인식의 정확도가 낮다고 판단할 수 있다. 이에 따라, 음성 인식 및 립 리딩을 조합하여 사용자의 명령을 분석하는 경우, 음성 인식을 통해 획득한 결과에 가중치를 낮게 할당할 수 있으며, 립 리딩을 통해 획득한 결과는 상대적으로 높은 가중치를 할당할 수 있다.
- [0151] 도 11은 본 발명의 다양한 실시 예에 따른 어플리케이션 실행 중 입술의 움직임을 감지하여 사용자의 발화를 인식하고, 인식된 발화에 상응하는 기능을 수행하는 방법(1100)에 관한 흐름도이다.
- [0152] 다양한 실시 예에 따르면, 전자 장치(101)는 1110 동작에서, 어플리케이션 실행 중 지능형 에이전트 서비스를 활성화 하는 입력을 수신할 수 있다.
- [0153] 다양한 실시 예에 따르면, 전자 장치(101)는 1120 동작에서, 전자 장치(101)의 상황 정보를 확인할 수 있다.
- [0154] 다양한 실시 예에 따르면, 전자 장치(101)는 1130 동작에서, 확인된 상황 정보에 기반하여, 카메라를 통해, 사용자의 영상 정보를 획득할 수 있다.
- [0155] 다양한 실시 예에 따르면, 전자 장치(101)는 1140 동작에서, 획득한 영상 정보에 포함된 입술의 움직임을 감지하여 사용자의 발화를 인식할 수 있다.
- [0156] 다양한 실시 예에 따르면, 전자 장치(101)는 1150 동작에서, 인식된 발화에 상응하는 기능을 수행할 수 있다.
- [0157] 다양한 실시 예에 따른 전자 장치에 있어서, 카메라; 마이크; 디스플레이; 메모리; 및 프로세서를 포함하고, 상기 프로세서는, 적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하고, 상기 전자 장치의 상황 정보를 확인하고, 상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하도록 제어하고, 상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하고, 및 상기 인식된 발화에 상응하는 기능을 수행하도록 설정될 수 있다.
- [0158] 상기 프로세서는, 상기 마이크를 통해, 상기 전자 장치의 주변 소음을 획득하고, 및 상기 획득한 주변 소음에 관한 정보를 상기 상황 정보로 저장하도록 설정될 수 있다.
- [0159] 상기 프로세서는, 상기 주변 소음이 기 설정된 값 이상인 경우, 상기 카메라를 활성화 하도록 설정될 수 있다.
- [0160] 상기 프로세서는, 상기 실행 중인 적어도 하나의 어플리케이션의 종류 또는 실행 상태 중 적어도 하나를 상기 상황 정보로 저장하고, 및 상기 실행 중인 적어도 하나의 어플리케이션이 음악 또는 동영상을 재생 중인 경우, 상기 카메라를 활성화 하도록 설정될 수 있다.
- [0161] 상기 프로세서는, 상기 마이크를 통해, 음성 신호를 수신하도록 제어하고, 상기 음성 신호 및 상기 입술의 움직임의 조합에 상응하는 조합 발화를 인식하고, 및 상기 인식된 조합 발화에 상응하는 기능을 수행하도록 설정될 수 있다.
- [0162] 상기 프로세서는, 상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 상기 입술의 움직임을 제외한, 상기 음성 신호에 상응하는 음성 인식 기반 발화를 인식하고, 및 상기 인식된 음성 인식 기반 발화에 상응하는 기능을 수행하도록 설정될 수 있다.
- [0163] 상기 프로세서는, 상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 사용자에게 상기 입술의 움직임

을 인식하는데 실패하였음을 알리는 사용자 인터페이스를 상기 디스플레이를 통해 표시하도록 설정될 수 있다.

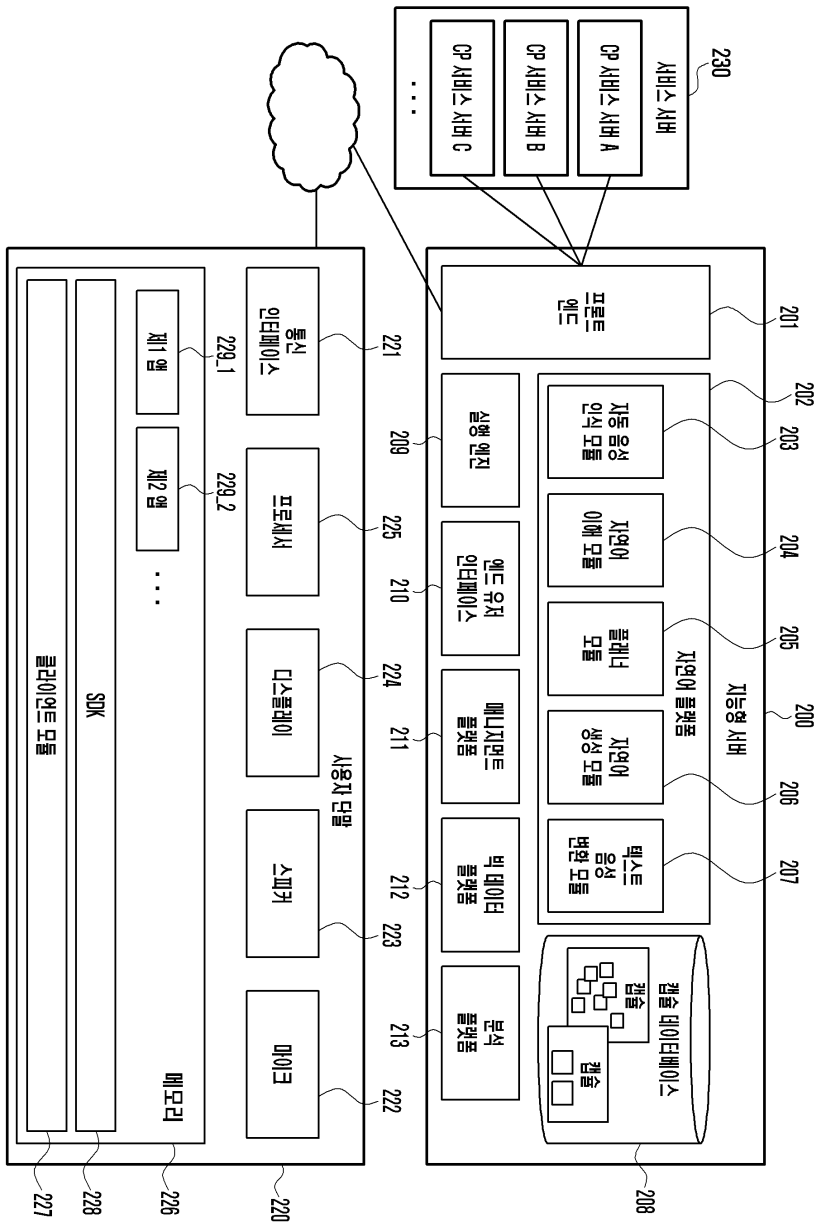
- [0164] 상기 프로세서는, 상기 획득한 영상 정보를 기반으로, 복수 개의 입술을 감지하는 경우, 상기 음성 신호 및 상기 복수 개의 입술의 움직임들을 확인하고, 및 상기 사용자에게 해당하는 입술을 다른 입술과 구분되게 하는 사용자 인터페이스를 상기 디스플레이를 통해 표시하도록 설정될 수 있다.
- [0165] 상기 프로세서는, 상기 사용자에게 해당하는 입술의 움직임을 기반으로 인식된 제1 발화의 가중치를 상기 마이크를 통해 획득한 음성 신호를 기반으로 인식된 제2 발화의 가중치보다 더 높게 설정하여, 상기 조합 발화를 인식하도록 설정될 수 있다.
- [0166] 상기 프로세서는, 상기 지능형 에이전트 서비스가 활성화 된 경우에도, 상기 적어도 하나의 어플리케이션에 관한 적어도 하나의 정보를 끊김 없이(seamless) 제공하도록 설정될 수 있다.
- [0167] 다양한 실시 예에 따른 전자 장치의 방법에 있어서, 적어도 하나의 어플리케이션을 실행 중에 지능형 에이전트 서비스를 활성화하는 입력을 사용자로부터 수신하는 동작; 상기 전자 장치의 상황 정보를 확인하는 동작; 상기 확인된 상황 정보에 기반하여, 상기 카메라를 통해, 상기 사용자의 영상 정보를 획득하는 동작; 상기 획득한 영상 정보에 포함된 상기 사용자의 입술의 움직임을 감지하여 상기 사용자의 발화를 인식하는 동작; 및 상기 인식된 발화에 상응하는 기능을 수행하는 동작을 포함할 수 있다.
- [0168] 마이크를 통해, 상기 전자 장치의 주변 소음을 획득하는 동작; 및 상기 획득한 주변 소음에 관한 정보를 상기 상황 정보로 저장하는 동작을 포함할 수 있다.
- [0169] 상기 주변 소음이 기 설정된 값 이상인 경우, 상기 카메라를 활성화 하는 동작을 포함할 수 있다.
- [0170] 상기 실행 중인 적어도 하나의 어플리케이션의 종류 또는 실행 상태 중 적어도 하나를 상기 상황 정보로 저장하는 동작; 및 상기 실행 중인 적어도 하나의 어플리케이션이 음악 또는 동영상을 재생 중인 경우, 상기 카메라를 활성화 하는 동작을 포함할 수 있다.
- [0171] 마이크를 통해, 음성 신호를 수신하는 동작; 상기 음성 신호 및 상기 입술의 움직임을 조합에 상응하는 조합 발화를 인식하는 동작; 및 상기 인식된 조합 발화에 상응하는 기능을 수행하는 동작을 포함할 수 있다.
- [0172] 상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 상기 입술의 움직임을 제외한, 상기 음성 신호에 상응하는 음성 인식 기반 발화를 인식하는 동작; 및 상기 인식된 음성 인식 기반 발화에 상응하는 기능을 수행하는 동작을 포함할 수 있다.
- [0173] 상기 획득한 영상 정보의 밝기가 기 설정된 값 미만인 경우, 사용자에게 상기 입술의 움직임을 인식하는데 실패하였음을 알리는 사용자 인터페이스를 디스플레이를 통해 표시하는 동작을 포함할 수 있다.
- [0174] 상기 획득한 영상 정보를 기반으로, 복수 개의 입술을 감지하는 경우, 상기 음성 신호 및 상기 복수 개의 입술의 움직임들을 확인하는 동작; 및 상기 사용자에게 해당하는 입술을 다른 입술과 구분되게 하는 사용자 인터페이스를 디스플레이를 통해 표시하는 동작을 포함할 수 있다.
- [0175] 상기 사용자에게 해당하는 입술의 움직임을 기반으로 인식된 제1 발화의 가중치를 상기 마이크를 통해 획득한 음성 신호를 기반으로 인식된 제2 발화의 가중치보다 더 높게 설정하여, 상기 조합 발화를 인식하는 동작을 포함할 수 있다.
- [0176] 상기 지능형 에이전트 서비스가 활성화 된 경우에도, 상기 적어도 하나의 어플리케이션에 관한 적어도 하나의 정보를 끊김 없이(seamless) 제공하는 동작을 포함할 수 있다.

도면

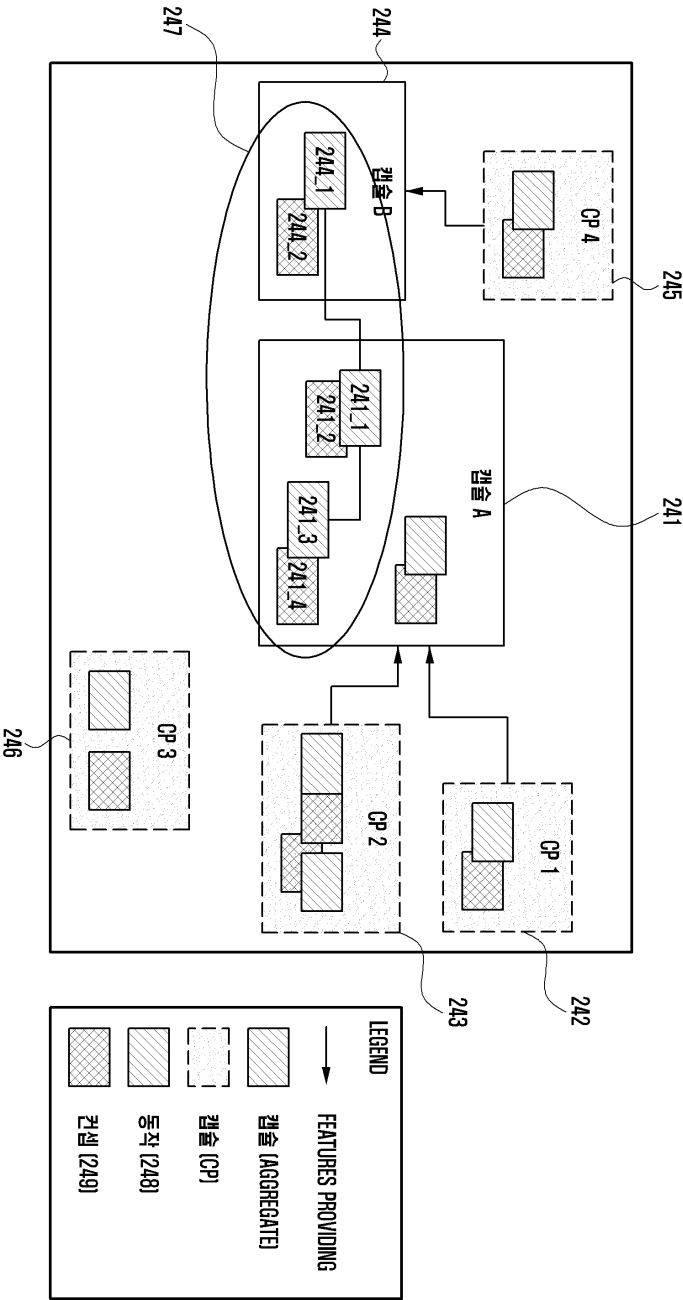
도면1



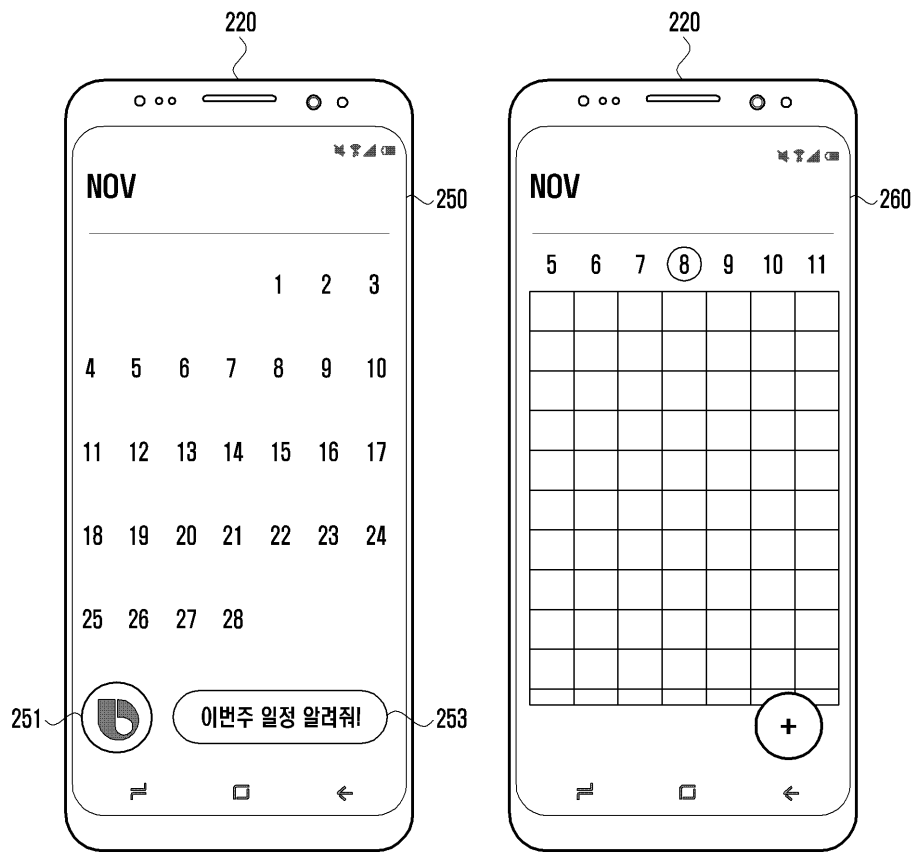
도면2a



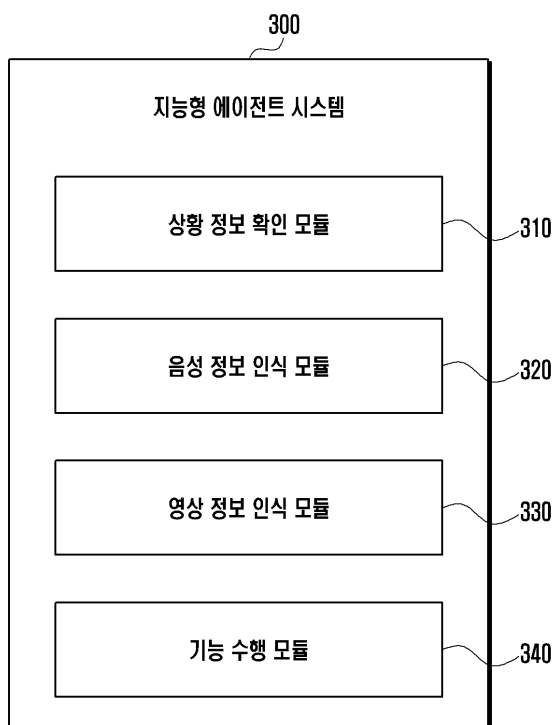
도면2b



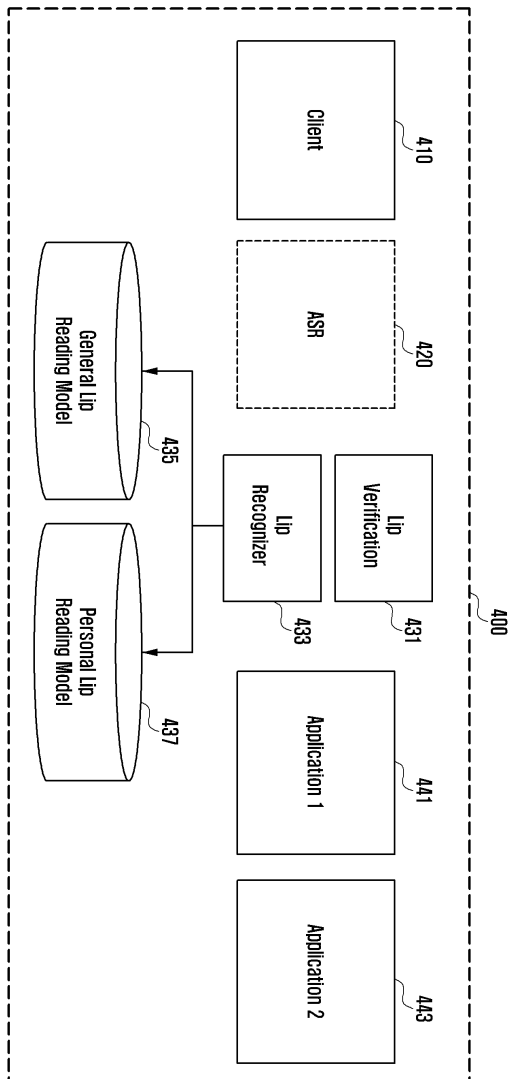
도면2c



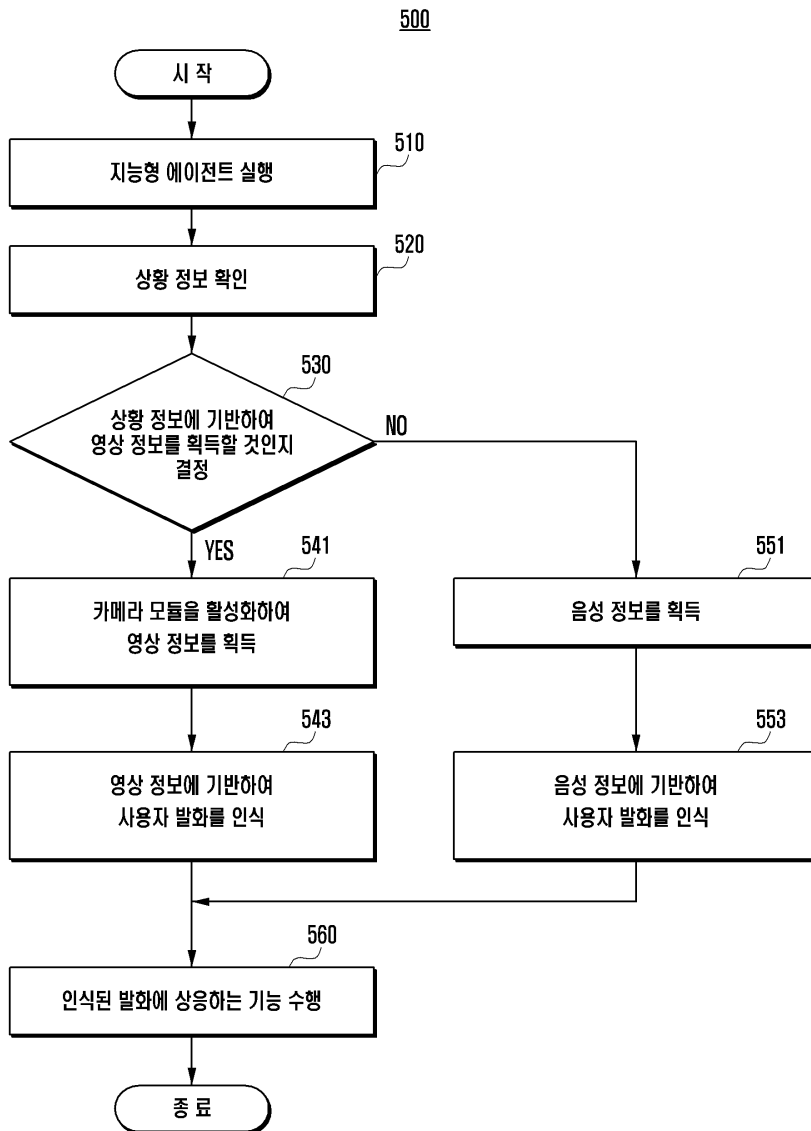
도면3



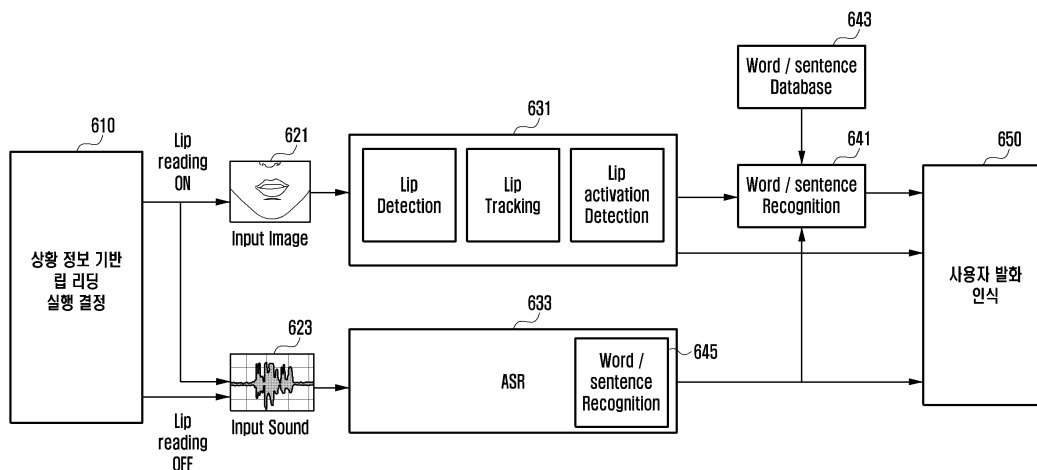
도면4



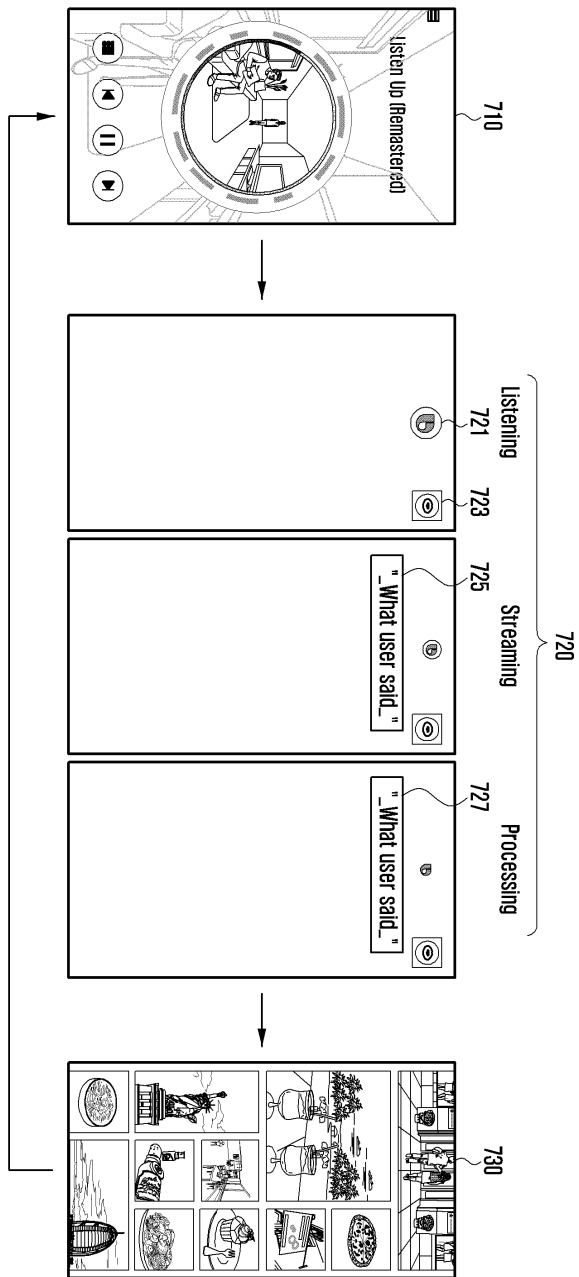
도면5



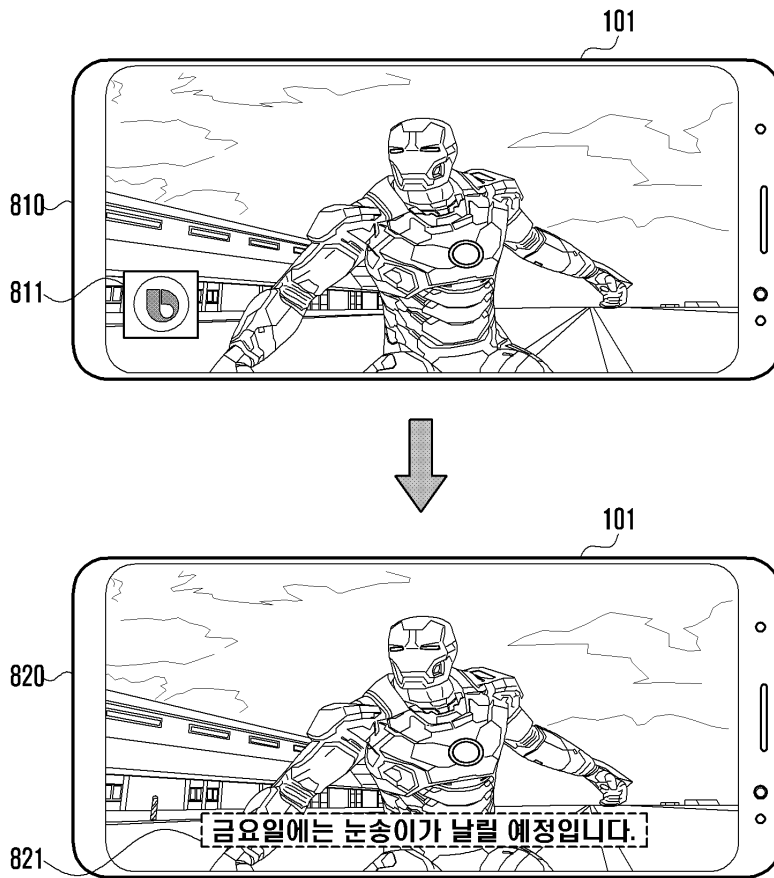
도면6



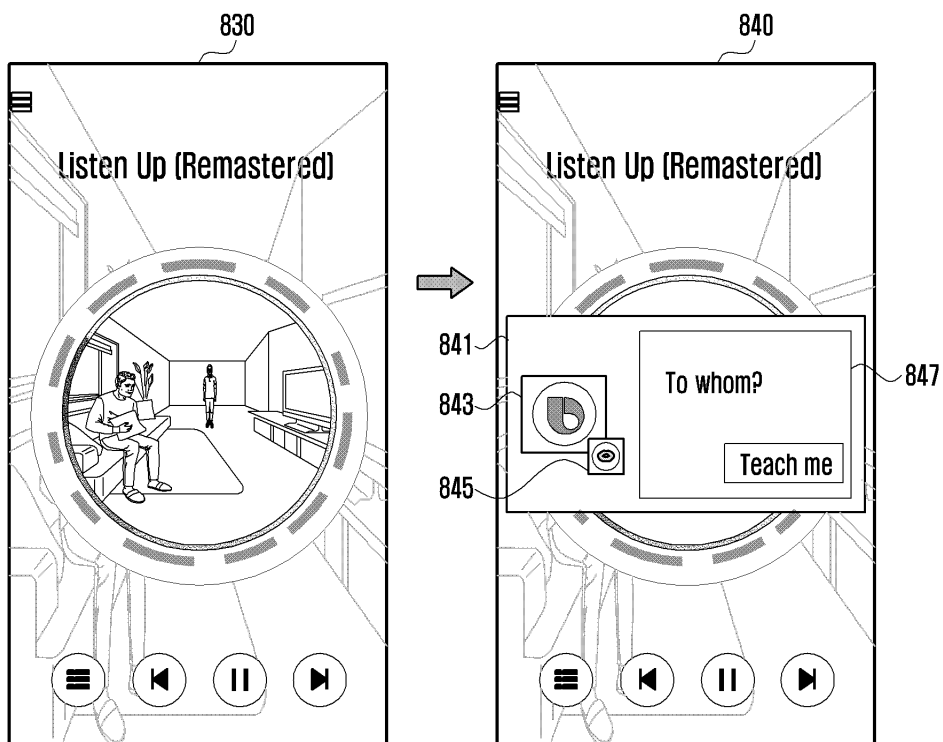
도면7



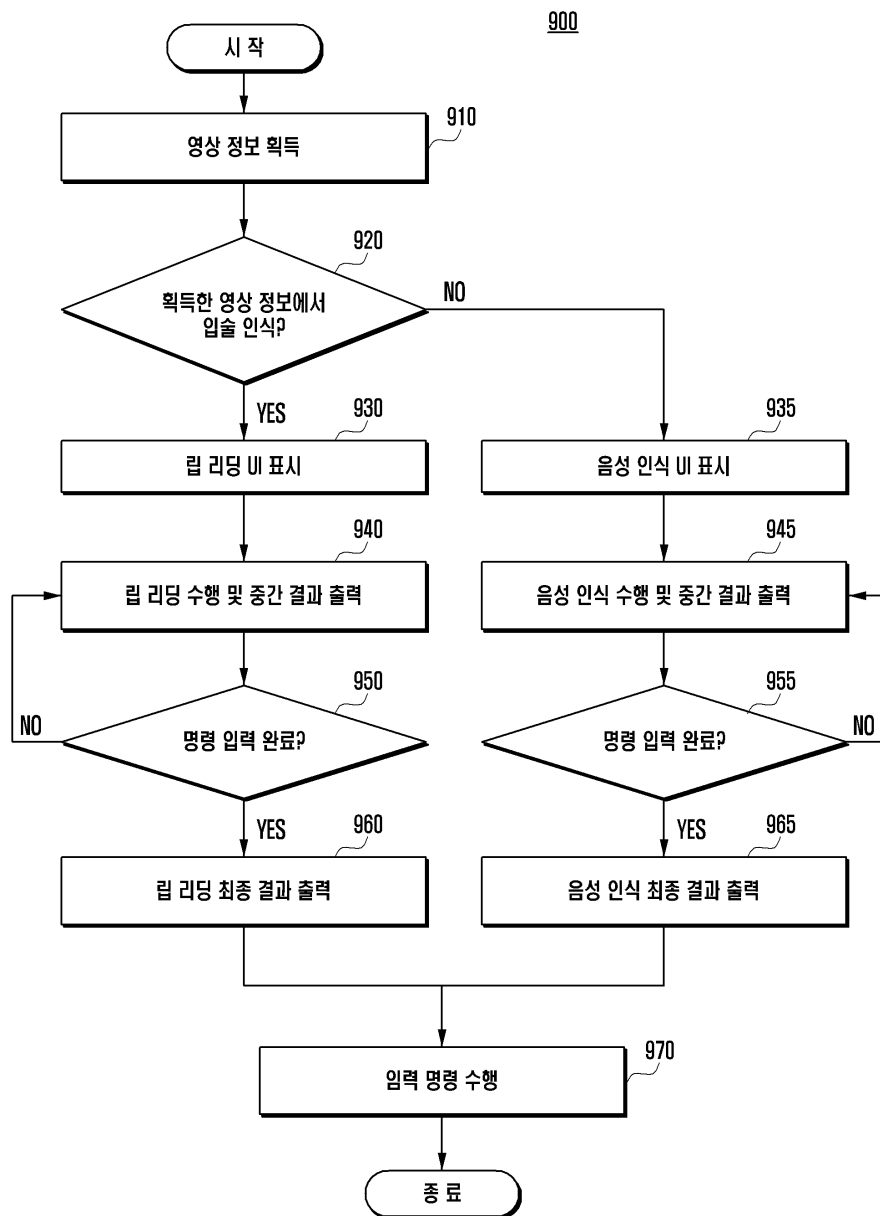
도면 8a



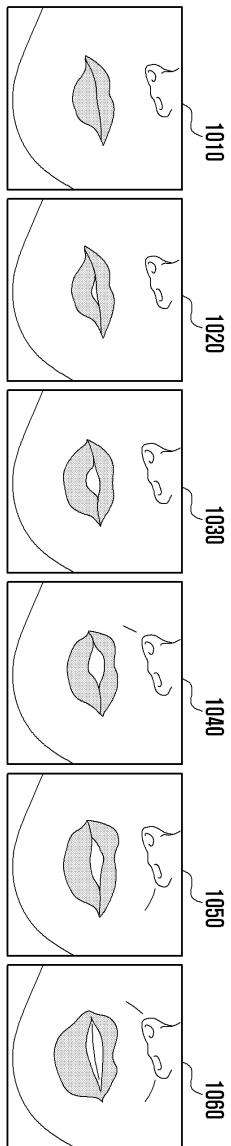
도면 8b



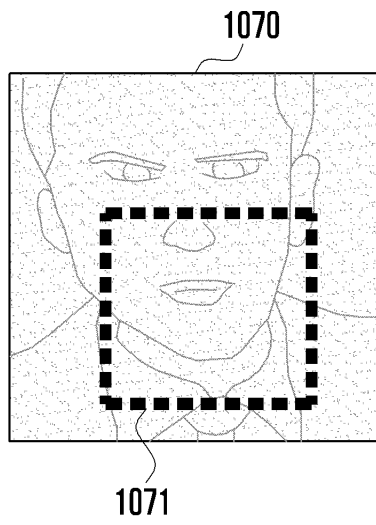
도면9



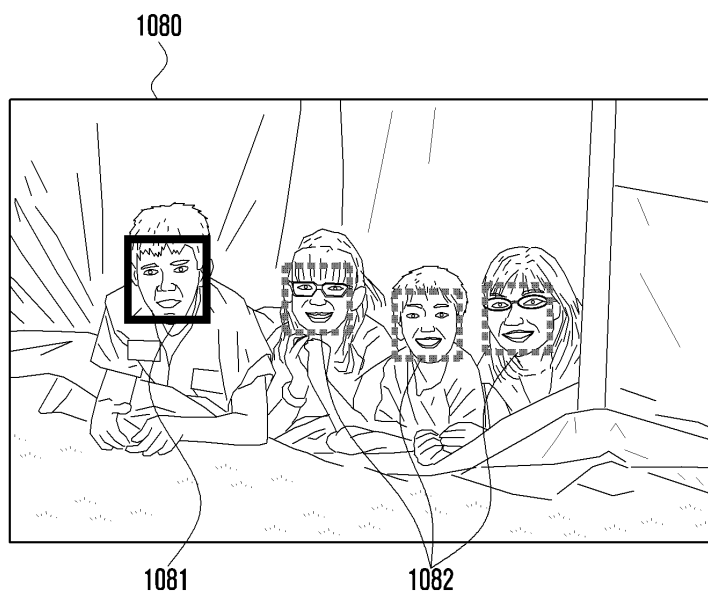
도면10a



도면10b



도면10c



도면11

