



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0059054
(43) 공개일자 2020년05월28일

(51) 국제특허분류(Int. Cl.)

G10L 15/22 (2006.01) G06F 3/16 (2018.01)
G10L 13/08 (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 3/16 (2019.02)

(21) 출원번호 10-2018-0143885

(22) 출원일자 2018년11월20일

심사청구일자 없음

(71) 출원인

삼성전자주식회사

경기도 수원시 영통구 삼성로 129 (매탄동)

(72) 발명자

이예슬

경기도 수원시 영통구 삼성로 129

김선옥

경기도 수원시 영통구 삼성로 129

(뒷면에 계속)

(74) 대리인

특허법인태평양

전체 청구항 수 : 총 20 항

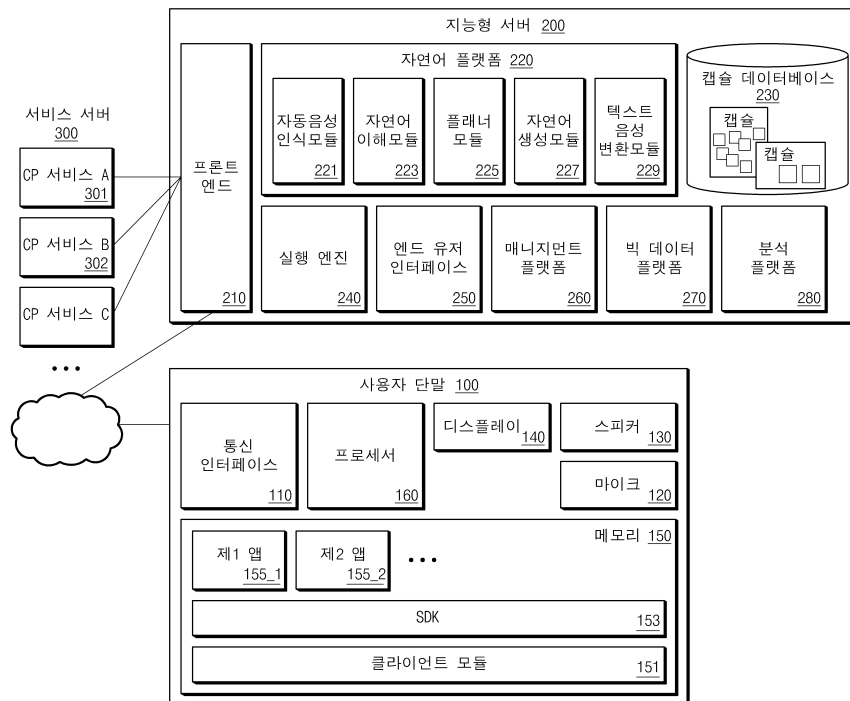
(54) 발명의 명칭 사용자 발화를 처리하는 전자 장치, 및 그 전자 장치의 제어 방법

(57) 요약

마이크, 스피커, 상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서, 및 상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고, 상기 메모리는, 자연어 이해(natural language understanding)(NLU) 모듈, 제1 응답 모델 및 제2 응답 모델을 저장하도록 구성되며, 실행될 때, 상기 프로세서

(뒷면에 계속)

대표도 - 도1



로 하여금, 제1 동작(operation)에서, 상기 마이크를 통해, 음성 기반의 지능형 어시스턴트 서비스를 호출하기 위한 제1 호출어를 포함하는 제1 음성 입력을 수신하고, 상기 제1 음성 입력에 기초하여 상기 제1 응답 모델을 선택하고, 상기 제1 음성 입력 이후에, 상기 마이크를 통해, 제2 음성 입력을 수신하고, 상기 NLU 모듈을 이용하여, 상기 제2 음성 입력을 처리하고, 상기 처리된 제2 음성 입력에 기초하여 제1 응답을 생성하고, 상기 제1 응답은 상기 제1 응답 모델을 이용하여 생성되고, 제2 동작(operation)에서, 상기 마이크를 통해, 상기 제1 호출어와 상이한 제2 호출어를 포함하는, 제3 음성 입력을 수신하고, 상기 제3 음성 입력에 기초하여 상기 제2 응답 모델을 선택하고, 상기 제3 음성 입력 이후에, 상기 마이크를 통해, 제4 음성 입력을 수신하고, 상기 NLU 모듈을 이용하여, 상기 제4 음성 입력을 처리하고, 상기 처리된 제4 음성 입력에 기초하여 제 2 응답을 생성하고, 상기 제2 응답은 상기 제2 응답 모델을 이용하여 생성되도록 하는 인스트럭션들을 저장한 시스템이 개시된다. 이 외에도 명세서를 통해 파악되는 다양한 실시 예가 가능하다.

(52) CPC특허분류

G10L 13/08 (2013.01)

G10L 15/04 (2013.01)

G10L 15/183 (2013.01)

G10L 15/26 (2013.01)

G10L 2015/225 (2013.01)

(72) 발명자

우혜림

경기도 수원시 영통구 삼성로 129

김경태

경기도 수원시 영통구 삼성로 129

명세서

청구범위

청구항 1

시스템에 있어서,

마이크;

스피커;

상기 마이크 및 상기 스피커와 작동적으로 연결된 적어도 하나의 프로세서; 및

상기 프로세서와 작동적으로 연결된 적어도 하나의 메모리를 포함하고,

상기 메모리는, 자연어 이해(natural language understanding)(NLU) 모듈, 제1 응답 모델 및 제2 응답 모델을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금,

제1 동작(operation)에서,

상기 마이크를 통해, 음성 기반의 지능형 어시스턴트 서비스를 호출하기 위한 제1 호출어를 포함하는 제1 음성 입력을 수신하고,

상기 제1 음성 입력에 기초하여 상기 제1 응답 모델을 선택하고,

상기 제1 음성 입력 이후에, 상기 마이크를 통해, 제2 음성 입력을 수신하고,

상기 NLU 모듈을 이용하여, 상기 제2 음성 입력을 처리하고,

상기 처리된 제2 음성 입력에 기초하여 제1 응답을 생성하고, 상기 제1 응답은 상기 제1 응답 모델을 이용하여 생성되고,

제2 동작(operation)에서,

상기 마이크를 통해, 상기 제1 호출어와 상이한 제2 호출어를 포함하는, 제3 음성 입력을 수신하고,

상기 제3 음성 입력에 기초하여 상기 제2 응답 모델을 선택하고,

상기 제3 음성 입력 이후에, 상기 마이크를 통해, 제4 음성 입력을 수신하고,

상기 NLU 모듈을 이용하여, 상기 제4 음성 입력을 처리하고,

상기 처리된 제4 음성 입력에 기초하여 제 2 응답을 생성하고, 상기 제2 응답은 상기 제2 응답 모델을 이용하여 생성되도록 하는 인스트럭션들을 저장하는, 시스템.

청구항 2

청구항 1에 있어서,

상기 제1 응답 모델은, 제1 자연어 응답 생성 모델, 및 제1 텍스트 음성 변환(text-to-speech)(TTS) 모델을 포함하고,

상기 제2 응답 모델은, 상기 제1 자연어 응답 생성 모델과 상이한 제2 자연어 응답 생성 모델, 및 상기 제1 TTS 모델과 상이한 제2 TTS 모델을 포함하는, 시스템.

청구항 3

청구항 1에 있어서,

사용자 단말 및 서버를 포함하고,

상기 사용자 단말은 상기 마이크를 포함하고,

상기 서버는 상기 적어도 하나의 메모리 및 상기 적어도 하나의 프로세서를 포함하는 시스템.

청구항 4

청구항 3에 있어서,

상기 사용자 단말은, 상기 제1 내지 상기 제4 음성 입력을 수신하고, 상기 제1 응답 및 상기 제2 응답을 사용자에게 제공하고,

상기 서버는, 상기 제1 음성 입력 및 상기 제3 음성 입력에 기초하여 상기 제1 음성 입력 모델 및 상기 제2 음성 입력 모델을 선택하고, 상기 제1 응답 및 상기 제2 응답을 생성하도록 하는, 시스템.

청구항 5

시스템에 있어서,

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(automatic speech recognition)(ASR) 모듈 및 자연어 이해(natural language understanding)(NLU) 모듈, 및 복수의 응답 모델들을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금,

상기 마이크를 통해, 제1 음성 입력을 수신하고,

상기 제1 음성 입력에 대해 상기 ASR 모듈을 이용하여 제1 텍스트 데이터를 추출하고,

상기 추출된 제 1 텍스트 데이터에 적어도 일부 기초하여, 상기 응답 모델들 중 적어도 하나를 선택하는, 시스템.

청구항 6

청구항 5에 있어서,

상기 인스트럭션들은 상기 프로세서로 하여금,

상기 추출된 제 1 텍스트 데이터에 적어도 일부 기초하여, 상기 응답 모델들 중 적어도 하나를 포함하는 리스트를 사용자에게 제공하고,

상기 응답 모델 리스트에서 적어도 하나의 응답 모델을 선택하는 사용자 입력을 수신하고,

상기 수신된 사용자 입력에 따라 상기 적어도 하나의 응답 모델을 선택하는, 시스템.

청구항 7

청구항 5에 있어서,

상기 인스트럭션들은 상기 프로세서로 하여금,

상기 제1 음성 입력에 적어도 일부 기반하여, 호출어를 인식하기 위한 음성 인식 모델을 생성하고, 상기 호출어는 음성 기반의 지능형 어시스턴트 서비스를 호출하기 위한 것인, 시스템.

청구항 8

청구항 7에 있어서.

상기 인스트럭션들은 상기 프로세서로 하여금,

상기 제 1 음성 입력 수신 이후에, 상기 마이크를 통해, 제 2 음성 입력을 수신하고,

상기 제1 음성 입력 및 상기 제2 음성 입력에 적어도 일부 기반하여, 상기 음성 인식 모델을 생성하도록 하는,

시스템.

청구항 9

청구항 5에 있어서,

상기 인스트럭션들은 상기 프로세서로 하여금,

상기 추출된 제1 텍스트 데이터와 상기 복수의 응답 모델들 사이의 관련성을 나타내는 값을 산출하고,

상기 복수의 응답 모델들 중 산출된 값이 지정된 값 이상인 적어도 하나의 응답 모델을 선택하도록 하는, 시스템.

청구항 10

청구항 5에 있어서,

상기 인스트럭션들은, 상기 적어도 하나의 프로세서로 하여금,

상기 추출된 제1 텍스트 데이터를 대신하여, 사용자 정보에 기초하여 상기 적어도 하나의 응답 모델을 선택하도록 하는, 시스템

청구항 11

청구항 10에 있어서,

상기 사용자 정보는 성별, 나이, 거주 지역, 및 사용자 단말 사용 정보 중 적어도 하나를 포함하는, 시스템.

청구항 12

청구항 5에 있어서,

상기 인스트럭션들은 상기 프로세서로 하여금,

상기 마이크를 통해 제2 사용자 음성 입력을 수신하고,

상기 NLU 모듈을 이용하여, 상기 제2 음성 입력을 처리하고, 상기 처리된 제2 음성 입력에 기초하여 제1 응답을 생성하고, 상기 제1 응답은 상기 선택된 적어도 하나의 응답 모델을 이용하도록 하는, 시스템.

청구항 13

청구항 5에 있어서,

상기 복수의 응답 모델은 텍스트 음성 변환(text-to-speech)(TTS) 모듈을 포함하는, 시스템.

청구항 14

음성 입력을 처리하는 방법에 있어서,

제1 음성 입력을 수신하는 동작;

상기 제1 음성 입력에 대해 ASR 모듈을 이용하여 제1 텍스트 데이터를 추출하는 동작; 및

상기 추출된 제 1 텍스트 데이터에 적어도 일부 기초하여, 복수의 응답 모델들 중 적어도 하나를 선택하는 동작;을 포함하는, 방법.

청구항 15

청구항 14에 있어서,

상기 제1 음성 입력에 적어도 일부 기반하여, 호출어를 인식하기 위한 음성 인식 모델을 생성하는 방법;을 더 포함하고, 상기 호출어는 음성 기반의 지능형 어시스턴트 서비스를 호출하기 위한 것인, 방법.

청구항 16

청구항 15에 있어서,

상기 제 1 음성 입력 수신 이후에, 상기 마이크를 통해, 제 2 음성 입력을 수신하는 동작; 및

상기 제1 음성 입력 및 상기 제2 음성 입력에 적어도 일부 기반하여, 음성 인식 모델을 생성하는 동작;을 더 포함하는, 방법.

청구항 17

청구항 14에 있어서,

상기 응답 모델들 중 적어도 하나를 선택하는 동작은,

상기 추출된 제1 텍스트 데이터와 상기 복수의 응답 모델들 사이의 관련성을 나타내는 값을 산출하는 동작; 및

상기 복수의 응답 모델들 중 산출된 값이 지정된 값 이상인 적어도 하나의 응답 모델을 선택하도록 하는 동작;을 포함하는, 방법.

청구항 18

청구항 14에 있어서,

상기 응답 모델들 중 적어도 하나를 선택하는 동작은,

상기 추출된 제1 텍스트 데이터를 대신하여, 사용자 정보에 기초하여 상기 적어도 하나의 응답 모델을 선택하는 동작;을 포함하는, 방법.

청구항 19

청구항 18에 있어서,

상기 사용자 정보는 성별, 나이, 거주 지역, 및 사용자 단말 사용 정보 중 적어도 하나를 포함하는, 방법.

청구항 20

청구항 14에 있어서,

제2 사용자 음성 입력을 수신하는 동작; 및

NLU 모듈을 이용하여, 상기 제2 음성 입력을 처리하고, 상기 처리된 제2 음성 입력에 기초하여 제1 응답을 생성하고, 상기 제1 응답은 상기 선택된 적어도 하나의 응답 모델을 이용하는 동작을 포함하는, 방법.

발명의 설명

기술 분야

[0001] 본 문서에서 개시되는 실시 예들은, 사용자 발화를 처리하는 기술과 관련된다.

배경 기술

[0002] 키보드나 마우스를 이용한 전통적인 입력 방식에 부가하여, 최근의 전자 장치들은 음성 입력과 같은 다양한 입력 방식을 지원할 수 있다. 예를 들어, 스마트폰이나 태블릿과 같은 전자 장치들은 음성 인식 서비스가 실행된 상태에서 입력되는 사용자의 음성을 인식하고, 음성 입력에 대응되는 동작을 실행하거나 검색 결과를 제공할 수 있다.

[0003] 근래 음성 인식 서비스는 자연어를 처리하는 기술을 기반으로 발전하고 있다. 자연어를 처리하는 기술은 사용자 발화의 의도를 파악하고, 의도에 맞는 결과를 사용자에게 제공하는 기술이다.

발명의 내용

해결하려는 과제

[0004] 사용자 단말은 수신된 음성 입력에 포함된 호출어를 인식하면 단말의 음성 인식 서비스를 일반적인 음성 입력을 처리하기 위한 활성화 상태로 변경시킬 수 있다. 사용자는 사용자가 원하는 단어를 호출어로 설정할 수 있다. 이에 따라, 호출어는 사용자가 음성 입력을 통해 사용자 단말을 제어하기 위한 시작점일 뿐만 아니라 사용자와

관련된 취향 정보를 포함할 수 있다. 그러나 사용자 단말은 하드웨어적인 한계 때문에 개인에 의해 설정된 호출어의 의미를 인식하고, 인식된 정보를 이용하여 개인화된 서비스를 제공하기 어려울 수 있다.

- [0005] 본 발명의 다양한 실시 예에 따른 통합 지능(integrated intelligence) 시스템은 사용자에게 의해 설정된 호출어를 이용하여 사용자에게 개인화된 음성 응답 서비스를 제공하고자 한다.

과제의 해결 수단

- [0006] 본 문서에 개시되는 일 실시 예에 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, 자연어 이해(natural language understanding)(NLU) 모듈, 제1 응답 모델 및 제2 응답 모델을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 제1 동작(operation)에서, 상기 마이크를 통해, 음성 기반의 지능형 어시스턴트 서비스를 호출하기 위한 제1 호출어를 포함하는 제1 음성 입력을 수신하고, 상기 제1 음성 입력에 기초하여 상기 제1 응답 모델을 선택하고, 상기 제1 음성 입력 이후에, 상기 마이크를 통해, 제2 음성 입력을 수신하고, 상기 NLU 모듈을 이용하여, 상기 제2 음성 입력을 처리하고, 상기 처리된 제2 음성 입력에 기초하여 제1 응답을 생성하고, 상기 제1 응답은 상기 제1 응답 모델을 이용하여 생성되고, 제2 동작(operation)에서, 상기 마이크를 통해, 상기 제1 호출어와 상이한 제2 호출어를 포함하는, 제3 음성 입력을 수신하고, 상기 제3 음성 입력에 기초하여 상기 제2 응답 모델을 선택하고, 상기 제3 음성 입력 이후에, 상기 마이크를 통해, 제4 음성 입력을 수신하고, 상기 NLU 모듈을 이용하여, 상기 제4 음성 입력을 처리하고, 상기 처리된 제4 음성 입력에 기초하여 제2 응답을 생성하고, 상기 제2 응답은 상기 제2 응답 모델을 이용하여 생성되도록 하는 인스트럭션들을 저장할 수 있다.

- [0007] 또한, 본 문서에 개시되는 일 실시 예에 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(automatic speech recognition)(ASR) 모듈 및 자연어 이해(natural language understanding)(NLU) 모듈, 및 복수의 응답 모델들을 저장하도록 구성되며, 실행될 때, 상기 프로세서로 하여금, 상기 마이크를 통해, 제1 음성 입력을 수신하고, 상기 제1 음성 입력에 대해 상기 ASR 모듈을 이용하여 제1 텍스트 데이터를 추출하고, 상기 추출된 제1 텍스트 데이터에 적어도 일부 기초하여, 상기 응답 모델들 중 적어도 하나를 선택할 수 있다.

- [0008] 또한, 본 문서에 개시되는 일 실시 예에 따른 음성 입력을 처리하는 방법은, 제1 음성 입력을 수신하는 동작, 상기 제1 음성 입력에 대해 ASR 모듈을 이용하여 제1 텍스트 데이터를 추출하는 동작, 및 상기 추출된 제1 텍스트 데이터에 적어도 일부 기초하여, 복수의 응답 모델들 중 적어도 하나를 선택하는 동작을 포함할 수 있다.

발명의 효과

- [0009] 본 문서에 개시되는 실시 예들에 따르면, 통합 지능 시스템은 사용자가 설정 가능한 호출어에 기초하여 응답 모델을 선택하고, 선택된 응답 모델을 이용하여 음성 응답을 생성함으로써, 호출어를 설정한 사용자에게 대해 개인화된 음성 어시스턴트 서비스(또는, 음성 기반의 지능형 어시스턴트 서비스)를 제공할 수 있다. 이에 따라, 사람과 직접 대화하는 것과 같은 사용자 경험성, 및 사용자 단말에 대한 사용자의 소유감뿐만 아니라, 그에 따른 음성 어시스턴트 서비스의 사용성을 증가시킬 수 있다.

- [0010] 이 외에, 본 문서를 통해 직접적 또는 간접적으로 파악되는 다양한 효과들이 제공될 수 있다.

도면의 간단한 설명

- [0011] 도 1은 일 실시 예에 따른 통합 지능(integrated intelligence) 시스템을 나타낸 블록도이다.
- 도 2는 다양한 실시 예에 따른, 컨셉과 액션의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.
- 도 3은 일 실시예에 따라, 지능형 앱을 통해 수신된 음성 입력을 처리하는 화면을 표시하는 사용자 단말을 도시하는 도면이다.
- 도 4a는 다양한 실시 예에 따른 통합 지능 시스템의 구성을 나타낸 블록도이다.
- 도 4b는 다양한 실시 예에 따른 통합 지능 시스템의 동작을 나타낸 순서도이다.
- 도 5a는 일 실시 예에 따른 통합 지능 시스템이 호출어를 등록하는 방법을 나타낸 순서도이다.

도 5b는 다양한 실시 예에 따른 통합 지능 시스템의 동작을 나타낸 순서도이다.

도 6은 일 실시 예에 따른 사용자 단말이 호출어를 등록하는 화면을 나타낸 도면이다.

도 7은 일 실시 예에 따른 통합 지능 시스템이 응답 모델을 설정하는 방법을 나타낸 순서도이다.

도 8은 일 실시 예에 따른 통합 지능 시스템이 사용자 단말의 초기 응답 모델을 설정하는 방법을 나타낸 도면이다.

도 9a 및 도 9b는 일 실시 예에 따른 통합 지능 시스템이 사용자 단말의 초기 응답 모델을 설정하는 방법을 나타낸 도면들이다.

도 10a 및 도 10b는 일 실시 예에 따른 응답 모델이 통합 지능 시스템에 설정되는 방법을 나타낸 도면들이다.

도 11은 일 실시 예에 따른 지능형 서버가 응답 모델을 결정하는 방법을 나타낸 흐름도이다.

도 12는 일 실시 예에 따른 통합 지능 시스템이 음성 입력에 대한 응답을 제공하는 방법을 나타낸 흐름도이다.

도 13은 일 실시 예에 따른 지능형 서버의 텍스트 음성 변환 모듈의 동작을 나타낸 도면이다.

도 14a 및 14b는 일 실시 예에 따른 통합 지능 시스템이 선택된 응답 모델에 따른 서비스를 제공하는 것을 나타낸 도면들이다.

도 15는 다양한 실시 예에 따른 네트워크 환경 내의 전자 장치의 블록도를 나타낸다.

도면의 설명과 관련하여, 동일 또는 유사한 구성요소에 대해서는 동일 또는 유사한 참조 부호가 사용될 수 있다.

발명을 실시하기 위한 구체적인 내용

- [0012] 이하, 본 발명의 다양한 실시 예가 첨부된 도면을 참조하여 기재된다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 실시 예의 다양한 변경(modification), 균등물(equivalent), 및/또는 대체물(alternative)을 포함하는 것으로 이해되어야 한다.
- [0014] 도 1은 일 실시예에 따른 통합 지능(integrated intelligence) 시스템을 나타낸 블록도이다.
- [0015] 도 1을 참조하면, 일 실시예의 통합 지능 시스템은 사용자 단말(100), 지능형 서버(200), 및 서비스 서버(300)를 포함할 수 있다.
- [0016] 일 실시 예의 사용자 단말(100)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.
- [0017] 도시된 실시 예에 따르면, 사용자 단말(100)은 통신 인터페이스(110), 마이크(120), 스피커(130), 디스플레이(140), 메모리(150), 또는 프로세서(160)를 포함할 수 있다. 상기 열거된 구성요소들은 서로 작동적으로 또는 전기적으로 연결될 수 있다.
- [0018] 일 실시 예의 통신 인터페이스(110)는 외부 장치와 연결되어 데이터를 송수신하도록 구성될 수 있다. 일 실시 예의 마이크(120)는 소리(예: 사용자 발화)를 수신하여, 전기적 신호로 변환할 수 있다. 일 실시예의 스피커(130)는 전기적 신호를 소리(예: 음성)으로 출력할 수 있다. 일 실시 예의 디스플레이(140)는 이미지 또는 비디오를 표시하도록 구성될 수 있다. 일 실시 예의 디스플레이(140)는 또한 실행되는 앱(app)(또는, 어플리케이션 프로그램(application program))의 그래픽 사용자 인터페이스(graphic user interface)(GUI)를 표시할 수 있다.
- [0019] 일 실시 예의 메모리(150)는 클라이언트 모듈(151), SDK(software development kit)(153), 및 복수의 앱들(155)을 저장할 수 있다. 상기 클라이언트 모듈(151), 및 SDK(153)는 범용적인 기능을 수행하기 위한 프레임워크(framework)(또는, 솔루션 프로그램)를 구성할 수 있다. 또한, 클라이언트 모듈(151) 또는 SDK(153)는 음성 입력을 처리하기 위한 프레임워크를 구성할 수 있다.
- [0020] 일 실시 예의 메모리(150)는 상기 복수의 앱들(155)은 지정된 기능을 수행하기 위한 프로그램일 수 있다. 일 실

시 예에 따르면, 복수의 앱(155)은 제1 앱(155_1), 제2 앱(155_3)을 포함할 수 있다. 일 실시 예에 따르면, 복수의 앱(155) 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(155)은 프로세서(160)에 의해 실행되어 상기 복수의 동작들 중 적어도 일부를 순차적으로 실행할 수 있다.

[0021] 일 실시 예의 프로세서(160)는 사용자 단말(100)의 전반적인 동작을 제어할 수 있다. 예를 들어, 프로세서(160)는 통신 인터페이스(110), 마이크(120), 스피커(130), 및 디스플레이(140)와 전기적으로 연결되어 연결되어 지정된 동작을 수행할 수 있다.

[0022] 일 실시 예의 프로세서(160)는 또한 상기 메모리(150)에 저장된 프로그램을 실행시켜 지정된 기능을 수행할 수 있다. 예를 들어, 프로세서(160)는 클라이언트 모듈(151) 또는 SDK(153) 중 적어도 하나를 실행하여, 음성 입력을 처리하기 위한 이하의 동작을 수행할 수 있다. 프로세서(160)는, 예를 들어, SDK(153)를 통해 복수의 앱(155)의 동작을 제어할 수 있다. 클라이언트 모듈(151) 또는 SDK(153)의 동작으로 설명된 이하의 동작은 프로세서(160)의 실행에 의한 동작일 수 있다.

[0023] 일 실시 예의 클라이언트 모듈(151)은 음성 입력을 수신할 수 있다. 예를 들어, 클라이언트 모듈(151)은 마이크(120)를 통해 감지된 사용자 발화에 대응되는 음성 신호를 수신할 수 있다. 상기 클라이언트 모듈(151)은 수신된 음성 입력을 지능형 서버(200)로 송신할 수 있다. 클라이언트 모듈(151)은 수신된 음성 입력과 함께, 사용자 단말(100)의 상태 정보를 지능형 서버(200)로 송신할 수 있다. 상기 상태 정보는, 예를 들어, 앱의 실행 상태 정보일 수 있다.

[0024] 일 실시 예의 클라이언트 모듈(151)은 수신된 음성 입력에 대응되는 결과를 수신할 수 있다. 예를 들어, 클라이언트 모듈(151)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(151)은 상기 수신된 결과를 디스플레이(140)에 표시할 수 있다.

[0025] 일 실시 예의 클라이언트 모듈(151)은 수신된 음성 입력에 대응되는 플랜을 수신할 수 있다. 클라이언트 모듈(151)은 플랜에 따라 앱의 복수의 동작을 실행한 결과를 디스플레이(140)에 표시할 수 있다. 클라이언트 모듈(151)은, 예를 들어, 복수의 동작의 실행 결과를 순차적으로 디스플레이에 표시할 수 있다. 사용자 단말(100)은, 다른 예를 들어, 복수의 동작을 실행한 일부 결과(예: 마지막 동작의 결과)만을 디스플레이에 표시할 수 있다.

[0026] 일 실시 예에 따르면, 클라이언트 모듈(151)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(151)은 상기 요청에 대응하여 상기 필요한 정보를 지능형 서버(200)로 송신할 수 있다.

[0027] 일 실시 예의 클라이언트 모듈(151)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.

[0028] 일 실시 예의 클라이언트 모듈(151)은 음성 인식 모듈을 포함할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(151)은 상기 음성 인식 모듈을 통해 제한된 기능을 수행하는 음성 입력을 인식할 수 있다. 예를 들어, 클라이언트 모듈(151)은 지정된 입력(예: 웨이크 업!)을 통해 유기적인 동작을 수행하기 위한 음성 입력을 처리하기 위한 지능형 앱을 수행할 수 있다.

[0029] 일 실시 예의 지능형 서버(200)는 통신 망을 통해 사용자 단말(100)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.

[0030] 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.

- [0031] 일 실시 예의 지능형 서버(200)는 생성된 플랜에 따른 결과를 사용자 단말(100)로 송신하거나, 생성된 플랜을 사용자 단말(100)로 송신할 수 있다. 일 실시 예에 따르면, 사용자 단말(100)은 플랜에 따른 결과를 디스플레이에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(100)은 플랜에 따른 동작을 실행한 결과를 디스플레이에 표시할 수 있다.
- [0032] 일 실시 예의 지능형 서버(200)는 프론트 엔드(front end)(210), 자연어 플랫폼(natural language platform)(220), 캡슐 데이터베이스(capsule DB)(230), 실행 엔진(execution engine)(240), 엔드 유저 인터페이스(end user interface)(250), 매니지먼트 플랫폼(management platform)(260), 빅 데이터 플랫폼(big data platform)(270), 또는 분석 플랫폼(analytic platform)(280)을 포함할 수 있다.
- [0033] 일 실시 예의 프론트 엔드(210)는 사용자 단말(100)로부터 수신된 음성 입력을 수신할 수 있다. 프론트 엔드(210)는 상기 음성 입력에 대응되는 응답을 송신할 수 있다.
- [0034] 일 실시 예에 따르면, 자연어 플랫폼(220)은 자동 음성 인식 모듈(automatic speech recognition module)(ASR module)(221), 자연어 이해 모듈(natural language understanding module)(NLU module)(223), 플래너 모듈(planner module)(225), 자연어 생성 모듈(natural language generator module)(NLG module)(227) 또는 텍스트 음성 변환 모듈(text to speech module)(TTS module)(229)를 포함할 수 있다.
- [0035] 일 실시 예의 자동 음성 인식 모듈(221)은 사용자 단말(100)로부터 수신된 음성 입력을 텍스트 데이터로 변환할 수 있다. 일 실시 예의 자연어 이해 모듈(223)은 음성 입력의 텍스트 데이터를 이용하여 사용자의 의도를 파악할 수 있다. 예를 들어, 자연어 이해 모듈(223)은 문법적 분석(syntactic analyze) 또는 의미적 분석(semantic analyze)을 수행하여 사용자의 의도를 파악할 수 있다. 일 실시 예의 자연어 이해 모듈(223)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.
- [0036] 일 실시 예의 플래너 모듈(225)은 자연어 이해 모듈(223)에서 결정된 의도 및 파라미터를 이용하여 플랜을 생성할 수 있다. 일 실시 예에 따르면, 플래너 모듈(225)은 상기 결정된 의도에 기초하여 태스크를 수행하기 위해 필요한 복수의 도메인을 결정할 수 있다. 플래너 모듈(225)은 상기 의도에 기초하여 결정된 복수의 도메인 각각에 포함된 복수의 동작을 결정할 수 있다. 일 실시 예에 따르면, 플래너 모듈(225)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(225)은 상기 복수의 동작, 및 상기 복수의 컨셉 사이의 관계를 단계적(또는, 계층적)으로 결정할 수 있다. 예를 들어, 플래너 모듈(225)은 복수의 컨셉에 기초하여 사용자의 의도에 기초하여 결정된 복수의 동작의 실행 순서를 결정할 수 있다. 다시 말해, 플래너 모듈(225)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(225)은 복수의 동작, 및 복수의 컨셉 사이의 연관 정보(예: 온톨로지(ontology))가 포함된 플랜을 생성할 수 있다. 상기 플래너 모듈(225)은 컨셉과 동작의 관계들의 집합이 저장된 캡슐 데이터베이스(230)에 저장된 정보를 이용하여 플랜을 생성할 수 있다.
- [0037] 일 실시 예의 자연어 생성 모듈(227)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(229)은 텍스트 형태의 정보를 음성 형태의 정보로 변경할 수 있다.
- [0038] 일 실시 예에 따르면, 자연어 플랫폼(220)의 기능의 일부 기능 또는 전체 기능은 사용자 단말(100)에서도 구현 가능할 수 있다.
- [0039] 상기 캡슐 데이터베이스(230)는 복수의 도메인에 대응되는 복수의 컨셉과 동작들의 관계에 대한 정보를 저장할 수 있다. 일 실시 예에 따르면, 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(230)는 CAN(concept action network)의 형태로 복수의 캡슐을 저장할 수 있다. 일 실시 예에 따르면, 복수의 캡슐은 캡슐 데이터베이스(230)에 포함된 기능 저장소(function registry)에 저장될 수 있다.
- [0040] 상기 캡슐 데이터베이스(230)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(230)

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(230)는 사용자 단말(100)을 통해 출력되는 정보의 레이아웃(layout) 정보를 저장하는 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(230)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(230)은 사용자 단말(100) 내에도 구현이 가능할 수 있다.

[0041] 일 실시 예의 실행 엔진(240)은 상기 생성된 플랜을 이용하여 결과를 산출할 수 있다. 엔드 유저 인터페이스(250)는 산출된 결과를 사용자 단말(100)로 송신할 수 있다. 이에 따라, 사용자 단말(100)은 상기 결과를 수신하고, 상기 수신된 결과를 사용자에게 제공할 수 있다. 일 실시 예의 매니지먼트 플랫폼(260)은 지능형 서버(200)에서 이용되는 정보를 관리할 수 있다. 일 실시 예의 빅 데이터 플랫폼(270)은 사용자의 데이터를 수집할 수 있다. 일 실시 예의 분석 플랫폼(280)을 지능형 서버(200)의 QoS(quality of service)를 관리할 수 있다. 예를 들어, 분석 플랫폼(280)은 지능형 서버(200)의 구성 요소 및 처리 속도(또는, 효율성)를 관리할 수 있다.

[0042] 일 실시 예의 서비스 서버(300)는 사용자 단말(100)에 지정된 서비스(예: 음식 주문 또는 호텔 예약)를 제공할 수 있다. 일 실시 예에 따르면, 서비스 서버(300)는 제3 자에 의해 운영되는 서버일 수 있다. 일 실시 예의 서비스 서버(300)는 수신된 음성 입력에 대응되는 플랜을 생성하기 위한 정보를 지능형 서버(200)에 제공할 수 있다. 상기 제공된 정보는 캡슐 데이터베이스(230)에 저장될 수 있다. 또한, 서비스 서버(300)는 플랜에 따른 결과 정보를 지능형 서버(200)에 제공할 수 있다.

[0043] 위에 기술된 통합 지능 시스템(10)에서, 상기 사용자 단말(100)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.

[0044] 일 실시 예에서, 상기 사용자 단말(100)은 내부에 저장된 지능형 앱(또는, 음성 인식 앱)을 통해 음성 인식 서비스를 제공할 수 있다. 이 경우, 예를 들어, 사용자 단말(100)은 상기 마이크를 통해 수신된 사용자 발화(utterance) 또는 음성 입력(voice input)을 인식하고, 인식된 음성 입력에 대응되는 서비스를 사용자에게 제공할 수 있다.

[0045] 일 실시 예에서, 사용자 단말(100)은 수신된 음성 입력에 기초하여, 단독으로 또는 상기 지능형 서버 및/또는 서비스 서버와 함께 지정된 동작을 수행할 수 있다. 예를 들어, 사용자 단말(100)은 수신된 음성 입력에 대응되는 앱을 실행시키고, 실행된 앱을 통해 지정된 동작을 수행할 수 있다.

[0046] 일 실시 예에서, 사용자 단말(100)이 지능형 서버(200) 및/또는 서비스 서버와 함께 서비스를 제공하는 경우에는, 상기 사용자 단말은, 상기 마이크(120)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(110)를 이용하여 지능형 서버(200)로 송신할 수 있다.

[0047] 일 실시 예에 따른 지능형 서버(200)는 사용자 단말(100)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.

[0048] 일 실시 예의 사용자 단말(100)은, 통신 인터페이스(110)를 이용하여 상기 응답을 수신할 수 있다. 사용자 단말

(100)은 상기 스피커(130)를 이용하여 사용자 단말(100) 내부에서 생성된 음성 신호를 외부로 출력하거나, 디스플레이(140)를 이용하여 사용자 단말(100) 내부에서 생성된 이미지를 외부로 출력할 수 있다.

[0049] 도 2는 다양한 실시 예에 따른, 컨셉과 동작의 관계 정보가 데이터베이스에 저장된 형태를 나타낸 도면이다.

[0050] 상기 지능형 서버(200)의 캡슐 데이터베이스(예: 캡슐 데이터베이스(230))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.

[0051] 상기 캡슐 데이터베이스는 복수의 도메인(예: 어플리케이션) 각각에 대응되는 복수의 캡슐(capsule(A)(401), capsule(B)(404))을 저장할 수 있다. 일 실시 예에 따르면, 하나의 캡슐(예: capsule(A)(401))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(402) 또는 CP 2 (403))가 대응될 수 있다. 일 실시 예에 따르면, 하나의 캡슐은 지정된 기능을 수행하기 위한 적어도 하나 이상의 동작(410) 및 적어도 하나 이상의 컨셉(420)을 포함할 수 있다.

[0052] 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(225)은 캡슐 데이터베이스에 저장된 캡슐을 이용하여 플랜을 생성할 수 있다. 예를 들어, 캡슐 A (410)의 동작들(4011, 4013)과 컨셉들(4012, 4014) 및 캡슐 B(404)의 동작(4041)과 컨셉(4042)를 이용하여 플랜(407)을 생성할 수 있다.

[0053] 도 3는 다양한 실시 예에 따른 사용자 단말이 지능형 앱을 통해 수신된 음성 입력을 처리하는 화면을 나타낸 도면이다.

[0054] 사용자 단말(100)은 지능형 서버(200)를 통해 사용자 입력을 처리하기 위해 지능형 앱을 실행할 수 있다.

[0055] 일 실시 예에 따르면, 310 화면에서, 사용자 단말(100)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(100)은, 예를 들어, 스케줄 앱을 실행한 상태에서 지능형 앱을 실행할 수 있다. 일 실시 예에 따르면, 사용자 단말(100)은 지능형 앱에 대응되는 오브젝트(예: 아이콘)(311)를 디스플레이(140)에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(100)은 사용자 발화에 의한 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(100)은 "이번주 일정 알려줘!"라는 음성 입력을 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말(100)은 수신된 음성 입력의 텍스트 데이터가 표시된 지능형 앱의 UI(user interface)(313)(예: 입력창)를 디스플레이에 표시할 수 있다.

[0056] 일 실시 예에 따르면, 320 화면에서, 사용자 단말(100)은 수신된 음성 입력에 대응되는 결과를 디스플레이에 표시할 수 있다. 예를 들어, 사용자 단말(100)은 수신된 사용자 입력에 대응되는 플랜을 수신하고, 플랜에 따라 '이번주 일정'을 디스플레이에 표시할 수 있다.

[0057] 도 4a는 다양한 실시 예에 따른 통합 지능 시스템(500)의 구성을 나타낸 블록도이다.

[0058] 도 4a를 참조하면, 통합 지능 시스템(500)은 사용자 단말(510), 및 지능형 서버(520)를 포함할 수 있다. 사용자 단말(510), 및 지능형 서버(520)는 도 1의 사용자 단말(100), 및 지능형 서버(200)와 유사할 수 있다.

[0059] 일 실시 예에 따르면, 사용자 단말(510)은 메모리(511)(예: 도 1의 메모리(150)), 서버 프로세서(513), 및 메인 프로세서(515)(예: 도 1의 프로세서(160))를 포함할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)의 구성은 이에 한정되지 않고, 도 1의 사용자 단말(100)의 구성을 더 포함할 수 있다.

[0060] 일 실시 예에 따르면, 메모리(511)는 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 테마(theme) 선택 모듈(511d), 및 호출어 인식 모델 데이터베이스(DB)(511e)를 저장할 수 있다. 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 테마 선택 모듈(511d)은 범용적인 기능을 수행하기 위한 프레임워크일 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 및 테마 선택 모듈(511d)은 프로세서(예: 서버 프로세서(513) 및 메인 프로세서(515))에 의해 실행되어 그 기능이 구현될 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 및 테마 선택 모듈(511d)은 소프트웨어뿐만 아니라 하드웨어로도 구현될 수 있다.

[0061] 일 실시 예에 따르면, 호출어 인식 모듈(511a)이 인식하는 호출어를 포함하는 사용자 발화는 음성 인식 서비스

(또는, 음성 기반의 지능형 어시스턴트 서비스)를 호출하는 웨이크업 발화(wake-up utterance)일 수 있다. 예를 들어, 호출어를 포함하는 음성 입력은 일반적인 음성 입력을 처리할 수 있는 상태로 변경하기 위한 음성 입력일 수 있다. 상기 일반적인 음성 입력은, 예를 들어, 지정된 기능(예: 메시지 송수신)을 수행하기 위한 음성 입력일 수 있다.

[0062] 일 실시 예에 따르면, 메모리(511)는 적어도 하나의 메모리를 포함할 수 있다. 예를 들어, 메모리(511)는 호출어 인식 모델 데이터베이스(511e)를 저장하기 위한 별도의 메모리를 포함할 수 있다. 다시 말해, 메모리(511)는 프로세서(예: 서브 프로세서(513) 및 메인 프로세서(515))의 동작을 제어하기 위한 명령어를 저장하는 제1 메모리, 및 호출어 인식 모델 데이터베이스(511e)를 저장하는 제2 메모리를 포함할 수 있다. 상기 제1 메모리는, 예를 들어, 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 및 테마 선택 모듈(511d)을 저장할 수 있다. 일 실시 예에 따르면, 상기 제2 메모리는 상기 제1 메모리와 물리적으로 분리된 메모리일 수 있다. 상기 제2 메모리는 사용자 단말(510)에서 제공하는 음성 인식 서비스 또는 음성 기반의 지능형 어시스턴트 서비스가 활성화되기 전의 상태에서 서브 프로세서(513)에 의해 접근 가능할 수 있다. 서브 프로세서(513)는, 예를 들어, 사용자 단말(510)에서 제공하는 음성 인식 서비스 또는 음성 기반의 지능형 어시스턴트 서비스를 활성화시키기 위한 음성 입력을 인식하기 위해 제2 메모리에 저장된 정보(예: 호출어 인식 모델 정보)를 읽어올 수 있다. 다른 예를 들어, 메모리(511)는 호출어 인식 모듈(511a), 클라이언트 모듈(511b), 호출어 훈련 모듈(511c), 테마 선택 모듈(511d), 및 호출어 인식 모델 데이터베이스(511e)를 저장하기 위한 하나의 메모리를 포함할 수 있다. 다시 말해, 메모리(511)는 호출어 인식 모델 데이터베이스(511e)를 위한 메모리를 별도로 포함하지 않을 수 있다.

[0063] 일 실시 예에 따르면, 서브 프로세서(513)는 사용자 단말(510)의 일부 동작을 제어할 수 있다. 다시 말해, 서브 프로세서(513)는 사용자 단말(510)을 제한적으로 제어할 수 있다. 예를 들어, 서브 프로세서(513)는 지정된 단어(또는, 호출어)를 인식하여 메인 프로세서(515)를 활성화시킬 수 있다. 다시 말해, 사용자 단말(510)은 서브 프로세서(513)에 의해 일부 동작(예: 시스템 부팅 상태 유지)만을 수행하는 비활성화 상태(또는, 대기 상태, 슬립 상태(sleep state))에서 복수의 서비스(예: 메시지 서비스, 전화 서비스)를 제공하기 위한 동작을 실행하는 활성화 상태로 변경될 수 있다. 일 실시 예에 따르면, 서브 프로세서(513)는 저전력을 소모하는 프로세서일 수 있다.

[0064] 일 실시 예에 따르면, 서브 프로세서(513)는 호출어 인식 모듈(511a)을 실행하여, 호출어를 인식하는 동작을 수행할 수 있다. 호출어 인식 모듈(511a)의 동작으로 설명된 이하의 동작은 서브 프로세서(513)의 실행에 의한 동작일 수 있다.

[0065] 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 제한된 개수의 단어를 인식할 수 있다. 예를 들어, 호출어 인식 모듈(511a)은 메인 프로세서(515)를 활성화시키기 위한 호출어를 인식할 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 호출어 인식 모델을 이용할 수 있다. 상기 호출어 인식 모델은 호출어를 인식하기 위해 필요한 정보를 포함할 수 있다. 예를 들어, 음성 인식 기능이 HMM 알고리즘(hidden Markov model algorithm)을 기반으로 수행되는 경우, 호출어 인식 모델은 상태 초기 확률(state initial probability), 상태 전의 확률(state transition probability), 관측 확률(observation probability) 등을 포함할 수 있다. 다른 예를 들어, 음성 인식 기능이 신경망 알고리즘(neural network algorithm)을 기반으로 수행되는 경우, 호출어 인식 모듈은 레이어(layer), 노드(node) 형태와 구조, 노드 별 가중치, 네트워크 연결 정보, 비선형 활성화 함수(activation function) 등의 신경망 모델 정보를 포함할 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)을 통해 호출어를 인식한 경우, 서브 프로세서(513)는 메인 프로세서(515)를 활성화시킬 수 있다.

[0066] 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 복수의 음성 입력의 유사도를 측정하여, 복수의 음성 입력이 동일한지 판단할 수 있다. 예를 들어, 호출어 인식 모듈(511a)은 복수의 음성 입력의 특징 벡터를 추출하고, DTW(dynamic time warping)를 이용하여 추출된 특징 벡터(feature vector)의 유사도를 측정할 수 있다. 다른 예를 들어, 호출어 인식 모듈(511a)은 음성 인식 기능이 HMM 알고리즘(hidden Markov model algorithm)을 기반으로 수행되는 경우, 포워드 백워드 확률(forward-backward probability)을 이용하여 복수의 음성 입력의 유사도를 측정할 수 있다. 또 다른 예를 들어, 호출어 인식 모듈(511a)은 신경망 알고리즘(neural network algorithm)을 기반으로 수행되는 경우, 음소 인식기를 통해 측정된 음소 레벨(phoneme level)을 이용하여 복수의 음성 입력의 유사도를 측정할 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 상기 측정된 유사도가 지정된 값 이상인 경우, 복수의 음성 입력을 동일한 것으로 판단할 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 호출어 훈련을 위한 복수의 음성 입력이 동일한 단어를 포함하였는지 판단할 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)의 호출어 훈련을 위한 동작은 메인 프로세서(515)에 의해 실행될 수

있다.

- [0067] 일 실시 예에 따르면, 메인 프로세서(515)는 클라이언트 모듈(511b)을 실행하여, 음성 입력을 처리하기 위한 동작을 수행할 수 있다. 상기 음성 입력은 지정된 태스크(task)를 수행하도록 명령하기 위한 사용자 입력일 수 있다. 클라이언트 모듈(511b)의 동작으로 설명된 이하의 동작은 메인 프로세서(515)의 실행에 의한 동작일 수 있다.
- [0068] 일 실시 예에 따르면, 클라이언트 모듈(511b)은 지정된 태스크를 수행하기 위한 음성 입력을 지능형 서버(520)로 송신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(511b)은 지능형 서버(520)를 통해 상기 음성 입력에 대응되는 플랜을 수신할 수 있다. 상기 플랜은 상기 지정된 태스크를 수행하기 위한 동작 정보를 포함할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(511b)은 상기 수신된 플랜에 따라 앱의 동작을 실행하여 상기 지정된 태스크를 수행함으로써, 그 결과를 사용자에게 제공할 수 있다.
- [0069] 일 실시 예에 따르면, 클라이언트 모듈(511b)은 호출어 등록을 위한 음성 입력을 수신하였을 때, 음성 입력을 지능형 서버(520)로 송신할 수 있다. 예를 들어, 클라이언트 모듈(511b)은 상기 음성 입력을 통신 인터페이스(예: 도 1의 통신 인터페이스(110))를 통해 지능형 서버(520)으로 송신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(511b)은 지능형 서버(520)로부터 음성 입력이 지정된 단어를 포함하고 있는지 여부를 확인한 결과를 수신할 수 있다. 예를 들어, 클라이언트 모듈(511b)은 상기 확인 결과를 통신 인터페이스를 통해 수신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(511b)은 상기 확인 결과에 기초하여, 호출어 등록 가능 여부를 판단할 수 있다. 이에 따라, 클라이언트 모듈(511b)은 수신된 음성 입력이 호출어로 등록 가능한 경우, 호출어 훈련 모듈(511c)을 통해 상기 수신된 음성 입력에 기초하여 호출어를 등록할 수 있다.
- [0070] 일 실시 예에 따르면, 메인 프로세서(515)는 호출어 훈련 모듈(511c)을 실행하여, 호출어를 등록(또는, 생성)하기 위한 동작을 수행할 수 있다. 사용자는 자신이 원하는 단어를 메인 프로세서(515)를 활성화시키기 위한 호출어로 등록할 수 있다. 호출어 훈련 모듈(511c)의 동작으로 설명된 이하의 동작은 메인 프로세서(515)의 실행에 의한 동작일 수 있다.
- [0071] 일 실시 예에 따르면, 호출어 훈련 모듈(511c)은 호출어를 등록시키기 위한 호출어 인식 훈련을 수행할 수 있다. 호출어 훈련 모듈(511c)은 반복적으로 수신된 음성 입력에 기초하여 호출어 인식 훈련을 수행할 수 있다. 예를 들어, 음성 인식 기능이 HMM 알고리즘을 기반으로 수행되는 경우, 호출어 훈련 모듈(511c)은 기대 값 최대화 알고리즘(expectation maximization(EM) algorithm)을 이용한 모델 훈련을 수행하거나, 최대 공산 선형 회귀법(maximum likelihood linear regression)(MLLR), 및 최대 귀납 추정법(maximum a posteriori estimation)(MAP)을 이용한 적응 훈련을 수행함으로써, 호출어 인식 모델을 생성할 수 있다. 다른 예를 들어, 음성 인식 기능이 신경망 알고리즘(neural network algorithm)을 기반으로 수행되는 경우, 호출어 훈련 모듈(511c)은 피드포워드(feedforward), 역방향 전파 알고리즘(backward-propagation algorithm)을 이용한 모델 훈련을 수행하거나, 선형 변환(linear transformation)을 이용한 적응 훈련을 수행할 수 있다. 일 실시 예에 따르면, 호출어 훈련 모듈(511c)은 호출어 인식 모듈(511a)을 통해 반복적으로 수신된 음성 입력에 동일한 호출어가 포함되어 있는지 판단할 수 있다. 이에 따라, 동일한 호출어를 포함한 음성 입력에 기초하여 호출어 인식 훈련을 수행할 수 있다. 일 실시 예에 따르면, 호출어 훈련 모듈(511c)은 호출어 인식 훈련을 통해 호출어 인식 모델을 생성할 수 있다.
- [0072] 일 실시 예에 따르면, 테마 선택 모듈(511d)은 응답 모델(예: 음성 응답 모델 또는 응답 테마 모델)을 선택하기 위한 리스트를 사용자에게 제공할 수 있다. 상기 리스트는, 예를 들어, 적어도 하나의 응답 모델을 포함할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 상기 리스트에 포함된 적어도 하나의 응답 모델을 선택할 수 있다. 예를 들어, 지능형 서버(520)는 호출어에 기초하여 상기 적어도 하나의 응답 모델을 결정할 수 있다. 다른 예를 들어, 지능형 서버(520)는 사용자 정보에 기초하여 상기 적어도 하나의 응답 모델을 선택할 수 있다. 일 실시 예에 따르면, 테마 선택 모듈(511d)은 상기 리스트에 포함된 적어도 하나의 응답 모델 중 적어도 하나를 선택하는 사용자 입력을 수신할 수 있다. 일 실시 예에 따르면, 테마 선택 모듈(511d)은 호출어 훈련 모듈(511c)을 통해 호출어 인식을 훈련할 때, 상기 리스트를 제공하고, 상기 리스트에 포함된 적어도 하나의 응답 모델을 선택하는 사용자 입력을 수신할 수 있다. 일 실시 예에 따르면, 클라이언트 모듈(511b)은 상기 수신된 사용자 입력을 지능형 서버(520)로 송신할 수 있다.
- [0073] 일 실시 예에 따르면, 지능형 서버(520)는 상기 수신된 사용자 입력에 따라 응답 모델을 선택하고, 상기 선택된 응답 모델을 이용하여 사용자 입력(예: 음성 입력)에 대응되는 응답을 생성할 수 있다. 예를 들어, 지능형 서버(520)는 상기 선택된 응답 모델에 따라 지정된 지정된 형식으로 사용자 입력에 대응되는 응답을 생성할 수

있다. 이에 따라, 사용자 단말(510)은 사용자의 선택에 의한 형식(또는, 테마)의 응답을 사용자에게 제공할 수 있다.

[0074] 일 실시 예에 따르면, 테마 선택 모듈(511d)은 호출어 훈련 모듈(511c)과 하나의 모듈로 구현될 수 있다. 다시 말해, 테마 선택 모듈(511d) 및 호출어 훈련 모듈(511c)은 하나의 하드웨어 모듈 또는 소프트웨어 모듈로 구현될 수 있다.

[0075] 일 실시 예에 따르면, 호출어 인식 모델 데이터베이스(511e)는 상기 생성된 호출어 인식 모델이 저장될 수 있다. 일 실시 예에 따르면, 호출어 인식 모듈(511a)은 호출어 인식 모델 데이터베이스(511e)에 저장된 호출어 인식 모델 정보를 이용하여 호출어를 인식할 수 있다.

[0076] 일 실시 예에 따르면, 지능형 서버(520)는 자동 음성 인식(ASR) 모듈(521)(예: 도 1의 자동 음성 인식 모듈(221)), 자연어 이해(NLU) 모듈(522)(예: 도 1의 자연어 이해 모듈(223)), 플래너 모듈(523)(예: 도 1의 플래너 모듈(225)), 자연어 응답 생성 모듈(524), 영상 응답 생성 모듈(525), 텍스트 음성 변환(TTS) 모듈(526), 대화 관리 모듈(527), 호출어 정제화 모듈(528), 호출명 데이터베이스(DB)(529a), 및 응답 모델 데이터베이스(DB)(529b)를 포함할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)의 구성은 이에 한정되지 않고, 도 1의 지능형 서버(200)의 구성을 더 포함할 수 있다. 지능형 서버(520)의 구성은 메모리에 저장된 프로그램이거나 데이터베이스일 수 있다.

[0077] 일 실시 예에 따르면, 자동 음성 인식 모듈(521)은 음성 입력을 텍스트 데이터로 변경할 수 있다. 일 실시 예에 따르면, 자동 음성 인식 모듈(521)은 HMM 알고리즘, wFST 알고리즘(weighted finite-state transducer algorithm), 신경망 알고리즘 등의 알고리즘 등을 이용하여 구현될 수 있다. 예를 들어, 자동 음성 인식 모듈(521)은 거리 측정(distance measure) 방법을 이용하여 지정된 단어와 음성 입력을 비교함으로써, 음성 입력을 텍스트 데이터로 변경할 수 있다. 상기 거리 측정 방법은, 예를 들어, Levenshtein 거리(distance), Jaro-Winkler 거리(distance) 등의 측정 방법을 포함할 수 있다. 상기 거리 측정 방법은, 다른 예를 들어, G2P(grapheme to phoneme)를 통해 발음열로 변환 후 음소 레벨(phoneme level)에서 텍스트 사이의 거리를 측정하는 방법을 포함할 수 있다.

[0078] 일 실시 예에 따르면, 자동 음성 인식 모듈(521)은 대어휘 연속어 음성인식 시스템(large vocabulary continuous speech recognition)(LVCSR)을 포함할 수 있다. 이에 따라, 자동 음성 인식 모듈(521)은 사용자 단말(510)의 호출어 인식 모듈(511a)보다 복잡한 계산 과정을 이용할 수 있고, 많은 단어 인식이 가능할 수 있다.

[0079] 일 실시 예에 따르면, 자연어 이해 모듈(522)은 자동 음성 인식 모듈(521)로부터 전달된 텍스트 데이터를 이용하여, 음성 입력에 대응되는 의도, 및 파라미터를 결정할 수 있다.

[0080] 일 실시 예에 따르면, 플래너 모듈(523)은 자연어 이해 모듈(522)에서 결정된 의도, 및 파라미터에 기초하여 음성 입력에 대응되는 플랜을 생성할 수 있다. 상기 플랜은 상기 음성 입력에 대응되는 태스크를 수행하기 위한 동작 정보를 포함할 수 있다. 일 실시 예에 따르면, 플래너 모듈(523)은 상기 태스크를 수행하기 위한 동작을 단계적으로 배열하고, 상기 배열된 동작의 실행에 입력되는 파라미터나, 실행에 의해 출력되는 결과 값을 정의한 컨셉을 결정함으로써, 플랜을 생성할 수 있다.

[0081] 일 실시 예에 따르면, 자연어 응답 생성 모듈(524)은 상기 결정된 의도, 및 파라미터에 기초하여 사용자에게 제공될 자연어 정보를 생성할 수 있다. 상기 자연어 정보는, 예를 들어, 시각 정보(예: 이미지 정보)뿐만 아니라, 청각 정보(예: 사운드 정보)를 포함할 수 있다. 일 실시 예에 따르면, 청각 정보는 사용자에게 음성으로 출력되는 정보일 수 있다. 상기 음성으로 제공될 청각 정보에 대응되는 텍스트는 지정된 템플릿(template)에 따라 생성될 수 있다. 상기 지정된 템플릿은, 예를 들어, 텍스트가 제공되는 형식과 관련된 것일 수 있다. 또한, 복잡한 텍스트는 자연어 생성 모듈(예: 도 1의 자연어 생성 모듈(227))을 통해 생성될 수 있다. 일 실시 예에 따르면, 자연어 응답 생성 모듈(524)은 복수의 템플릿 중 하나의 템플릿을 이용하여 자연어 정보를 생성할 수 있다. 일 실시 예에 따르면, 자연어 응답 생성 모듈(524)은 호출명을 이용하여 사용자에게 제공할 정보를 생성할 수 있다. 예를 들어, 자연어 응답 생성 모듈(524)은 호출명이 포함된 텍스트를 생성할 수 있다. 상기 호출명은, 예를 들어, 호출어를 정제하여 생성된 것일 수 있다. 상기 호출명은 호출명 데이터베이스(529a)에 저장될 수 있다.

[0082] 일 실시 예에 따르면, 영상 응답 생성 모듈(525)은 상기 결정된 의도, 및 파라미터에 기초하여 영상 정보를 생성할 수 있다. 상기 영상 정보는, 예를 들어, 자연어 응답 생성 모듈(524)을 통해 생성되는 자연어 정보 외의 정보로써, 이미지 정보 및 상기 이미지 정보에 대응되는 사운드 정보를 포함할 수 있다. 일 실시 예에 따르면,

상기 영상 정보는 지정된 템플릿에 따라 생성될 수 있다. 상기 지정된 템플릿은, 예를 들어, 이미지 오브젝트의 위치 정보, 색상 정보, 모양 정보 등을 포함할 수 있다. 일 실시 예에 따르면, 영상 응답 생성 모듈(525)은 복수의 템플릿 중 하나의 템플릿을 이용하여 영상 정보를 생성할 수 있다.

[0083] 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 자연어 응답 생성 모듈(524)에서 생성된 자연어 정보를 음성 신호로 변환할 수 있다. 텍스트 음성 변환 모듈(526)은 알고리즘에 따라 다양한 형식으로 구현될 수 있다. 예를 들어, 텍스트 음성 변환 모듈(526)은 유닛 접합 형식으로 구현될 수 있다. 상기 유닛 접합 형식으로 구현된 텍스트 음성 변환 모듈(526)은, 예를 들어, 대단위 음성 데이터베이스를 포함할 수 있다. 상기 대단위 음성 데이터베이스는 복수의 사람들을 통해 녹음된 음성 신호가 태깅된 텍스트 음성 변환 모델이 저장될 수 있다. 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 텍스트 음성 변환 모델을 이용하여 자연어 정보를 음성 신호로 변환할 수 있다. 다른 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 음성 요소를 조절하여 자연어 정보를 음성 신호로 변환할 수 있다. 예를 들어, 텍스트 음성 변환 모듈(526)은 피치(pitch), 발화 속도, 억양 등을 조절하여 자연어 정보를 음성 신호로 변환할 수 있다. 상기 음성 요소의 조절은 자연어 정보를 음성 신호로 변환하는 중에 조절(예: 파라메트릭 텍스트 음성 변환(parametric TTS) 방식)되거나, 음성 신호 변환 후에 조절(예: 후처리 방식)될 수 있다.

[0084] 일 실시 예에 따르면, 대화 관리 모듈(527)은 사용자와 대화의 흐름을 관리할 수 있다. 예를 들어, 대화 관리 모듈(527)은 사용자와의 대화 중 컨텍스트에 기초하여 수신된 음성 입력에 대한 응답을 결정할 수 있다. 일 실시 예에 따르면, 대화 관리 모듈(527)은 대화 관리 모델을 이용하여 음성 입력에 대한 응답을 결정할 수 있다. 상기 대화 관리 모델은, 예를 들어, 응답을 결정할 수 있다.

[0085] 일 실시 예에 따르면, 호출어 정제화 모듈(528)은 등록된 호출어를 정제하여 사용자 단말(510)의 호출명을 생성할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 상기 호출명을 이용하여 음성 입력에 대응되는 응답을 생성할 수 있다. 다시 말해, 지능형 서버(520)는 상기 호출명이 포함된 응답을 사용자 단말(510)에 제공할 수 있다. 상기 호출명은, 예를 들어, 사용자의 음성 입력에 대응되는 응답을 제공하는 음성 어시스턴트(the voice assistant)의 명칭(name)으로 사용될 수 있다. 일 실시 예에 따르면, 호출어 정제화 모듈(528)은 생성된 호출명을 호출명 데이터베이스(529a)에 저장할 수 있다.

[0086] 일 실시 예에 따르면, 지능형 서버(520)는 사용자 입력에 대응되는 응답을 제공하기 위한 응답 모델을 선택할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 호출어에 기초하여 응답 모델을 선택할 수 있다. 예를 들어, 지능형 서버(520)는 자동 음성 인식 모듈(521)을 통해 호출어에 대응되는 텍스트 데이터를 획득하고, 상기 획득된 텍스트 데이터에 기초하여 응답 모델을 선택할 수 있다. 다른 실시 예에 따르면, 지능형 서버(520)는 사용자 정보에 기초하여 응답 모델을 선택할 수 있다. 사용자 정보는, 예를 들어, 성별, 나이, 거주 지역, 및 사용자 단말 사용 정보 중 적어도 하나를 포함할 수 있다.

[0087] 일 실시 예에 따르면, 지능형 서버(520)는 선택된 응답 모델을 이용하여 수신된 음성 입력에 대응되는 응답을 생성할 수 있다. 예를 들어, 지능형 서버(520)는 선택된 응답 모델에 따른 형식으로 수신된 음성 입력에 대응되는 응답을 생성할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)의 응답을 생성하기 위한 구성은 선택된 음성 입력 모델에 기초하여 설정될 수 있다. 예를 들어, 자연어 응답 생성 모듈(524) 및 영상 응답 생성 모듈(525)은 선택된 음성 입력 모델에 대응되는 템플릿을 이용하여 사용자에게 제공되는 정보(예: 자연어 정보 및 영상 정보)를 생성하도록 설정될 수 있다. 다른 예를 들어, 텍스트 음성 변환 모듈(526)은 선택된 음성 입력 모델에 대응되는 텍스트 음성 변환 모델을 이용하여 자연어를 음성 신호로 변환할 수 있다. 대화 관리 모듈(527)은 선택된 음성 입력 모델에 대응되는 대화 관리 모델을 이용하여 응답을 결정할 수 있다.

[0088] 일 실시 예에 따르면, 응답 모델 데이터베이스(529b)는 음성 변환 모델의 정보를 저장할 수 있다. 상기 음성 변환 모델의 정보는, 예를 들어, 템플릿 정보, 텍스트 음성 변환 모델 정보, 및 대화 관리 모델 정보를 저장할 수 있다. 이에 따라, 지능형 서버(520)는 응답 모델 데이터베이스(529b)에 저장된 음성 변환 모델의 정보를 이용하여 수신된 음성 입력에 대응되는 응답을 생성할 수 있다. 일 실시 예에 따르면, 응답 모델 데이터베이스(529b)는 복수의 음성 변환 모델의 정보를 저장할 수 있다. 지능형 서버(520)는 등록된 호출어에 기초하여 복수의 음성 변환 모델 중 하나의 음성 변환 모델을 결정할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 복수의 호출어를 등록시킬 수 있다. 지능형 서버(520)는 인식된 호출어에 대응되는 음성 변환 모델을 이용하여 응답을 생성할 수 있다.

[0089] 일 실시 예에 따르면, 지능형 서버(520)는 사용자로부터 지정된 음성 입력을 수신하기 위한 안내 정보를 생성할 수 있다. 예를 들어, 지능형 서버(520)는 호출어 훈련을 위한 음성 입력을 수신하기 위한 안내 정보를 생성할

수 있다. 또한, 지능형 서버(520)는 부적절한 단어를 포함하는 음성 입력을 수신하였을 때, 상이한 음성 입력을 수신하기 위한 안내 정보를 생성할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 상기 생성된 안내 정보를 사용자 단말(510)으로 송신할 수 있다. 사용자 단말(510)은 상기 안내 정보를 출력할 수 있다.

- [0090] 도 4b는 다양한 실시 예에 따른 통합 지능 시스템(500)의 동작을 나타낸 순서도(530)이다.
- [0091] 일 실시 예에 따르면, 통합 지능 시스템(500)의 지능형 서버(520)는 상이한 호출어에 대해 상이한 응답 모델을 이용하여 응답을 제공할 수 있다.
- [0092] 일 실시 예에 따른 지능형 서버(520)는 동작 531 내지 동작 535를 포함하는 제1 동작(operation)을 수행할 수 있다.
- [0093] 동작 531에서, 지능형 서버(520)는 제1 호출어를 포함하는 제1 음성 입력을 수신할 수 있다. 제1 호출어는 지능형 어시스턴트 서비스를 호출할 수 있다. 제1 음성 입력은 마이크를 통해 수신할 수 있다.
- [0094] 동작 532에서, 지능형 서버(520)는 제1 음성 입력에 기초하여 제1 응답 모델을 선택할 수 있다.
- [0095] 동작 533에서, 지능형 서버(520)는 제2 음성 입력을 수신할 수 있다. 제2 음성 입력은 마이크를 통해 수신할 수 있다. 제2 음성 입력은 제1 음성 입력 이후에 수신할 수 있다. 제2 음성 입력은 제1 결과를 산출하기 위한 음성 입력일 수 있다.
- [0096] 동작 534에서, 지능형 서버(520)는 제2 음성 입력을 처리할 수 있다. 제2 음성 입력은 자연어 이해(NLU) 모듈(522)을 이용하여 처리할 수 있다.
- [0097] 동작 535에서, 지능형 서버(520)는 처리된 제2 음성 입력에 기초하여 제1 응답을 생성할 수 있다. 제1 응답은 제1 응답 모델을 이용하여 생성될 수 있다.
- [0098] 또한, 일 실시 예에 따른 지능형 서버(520)는 동작 536 내지 동작 540를 포함하는 제2 동작을 수행할 수 있다.
- [0099] 동작 536에서, 지능형 서버(520)는 제1 호출어와 상이한 제2 호출어를 포함하는 제3 음성 입력을 수신할 수 있다. 제3 음성 입력은 마이크를 통해 수신할 수 있다.
- [0100] 동작 537에서, 지능형 서버(520)는 제2 음성 입력에 기초하여 제2 응답 모델을 선택할 수 있다.
- [0101] 동작 538에서, 지능형 서버(520)는 제4 음성 입력을 수신할 수 있다. 제4 음성 입력은 제3 음성 입력 이후에 수신할 수 있다. 제4 음성 입력은 마이크를 통해 수신할 수 있다.
- [0102] 동작 539에서, 지능형 서버(520)는 제4 음성 입력을 처리할 수 있다. 제4 음성 입력은 자연어 이해(NLU) 모듈(522)을 이용하여 처리할 수 있다.
- [0103] 동작 540에서, 지능형 서버(520)는 처리된 제4 음성 입력에 기초하여 제2 응답을 생성할 수 있다. 제2 응답은 제2 응답 모델을 이용하여 생성될 수 있다.
- [0104] 도 5a는 일 실시 예에 따른 통합 지능 시스템(500)이 호출어를 등록하는 방법을 나타낸 순서도(600)이다.
- [0105] 도 5a를 참조하면, 사용자 단말(예: 도 4의 사용자 단말(510))은 지능형 서버(예: 도 4의 지능형 서버(520))를 통해 음성 입력에 부적절한 단어가 포함되어 있는지 확인하고, 호출어를 등록할 수 있다.
- [0106] 일 실시 예에 따르면, 610 동작에서, 사용자 단말은 호출어를 등록하기 위한 제1 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말은 “갤럭시”를 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말은 상기 수신된 제1 음성 입력을 지능형 서버로 송신할 수 있다.
- [0107] 일 실시 예에 따르면, 620 동작에서, 지능형 서버는 상기 수신된 제1 음성 입력을 인식할 수 있다. 예를 들어, 지능형 서버는 자동 음성 인식 모듈(예: 도 4의 자동 음성 인식 모듈(521))을 통해 제1 음성 입력을 텍스트 데이터(예: 갤럭시)로 변경하여, 제1 음성 입력을 인식할 수 있다.
- [0108] 일 실시 예에 따르면, 630 동작에서, 사용자 단말은 호출어 인식 훈련을 위한 추가 발화를 요청할 수 있다. 예를 들어, 사용자 단말은 제1 음성 입력(예: 갤럭시)과 동일한 제2 음성 입력을 수신하기 위한 제1 안내 정보(예: '갤럭시'라고 다시 말씀해주세요.)를 스피커, 또는 디스플레이를 통해 출력할 수 있다. 상기 제1 안내 정보는, 예를 들어, 제1 음성 입력을 이용하여 생성될 수 있다. 일 실시 예에 따르면, 사용자 단말은 지능형 서버로부터 상기 제1 안내 정보를 수신할 수 있다. 지능형 서버는 사용자 단말로부터 수신된 제1 음성 입력을 이용하여 상기 제1 안내 정보를 생성할 수 있다.

- [0109] 일 실시 예에 따르면, 640 동작에서, 사용자 단말은 상기 제1 음성 입력과 동일한 정보를 포함하는 제2 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말은 “갤럭시”를 다시 수신할 수 있다.
- [0110] 일 실시 예에 따르면, 650 동작에서, 사용자 단말은 상기 제1 음성 입력, 및 상기 제2 음성 입력에 기초하여 호출어를 인식하기 위한 호출어 인식 모델을 생성할 수 있다. 상기 호출어 인식 모델은, 예를 들어, HMM 알고리즘(hidden Markov model algorithm), 또는 신경망 알고리즘(neural network algorithm) 중 적어도 하나에 기초한 모델 훈련 또는 적응 훈련 알고리즘을 이용하여 생성될 수 있다. 일 실시 예에 따르면, 사용자 단말은 상기 생성된 호출어 인식 모델을 상기 메모리에 저장함으로써, 호출어를 등록시킬 수 있다.
- [0111] 일 실시 예에 따르면, 사용자 단말은, 640 동작에서, 상기 제1 음성 입력과 동일한 정보를 포함하는 음성 입력을 복수회(예: N회) 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말은, 650 동작에서, 상기 복수회의 음성 입력에 기초하여 호출어 인식 모델을 생성할 수 있다. 이에 따라, 사용자 단말은 호출어를 정확하게 인식할 수 있는 호출어 인식 모델을 생성할 수 있다.
- [0112] 이에 따라, 사용자 단말은 대기 상태에서 등록된 호출어를 포함하는 음성 입력을 수신하였을 때, 호출어를 인식하고, 음성 인식 서비스(또는, 음성 기반의 지능형 어시스턴트 서비스)를 활성화 상태로 변경될 수 있다.
- [0113] 도 5b는 다양한 실시 예에 따른 통합 지능 시스템(500)의 동작을 나타낸 순서도(660)이다.
- [0114] 동작 661에서, 지능형 서버(520)는 제1 음성 입력에 대해 자동 음성 인식(ASR) 모듈(521)을 이용하여 제1 텍스트 데이터를 추출할 수 있다. 제1 음성 입력은 동작 610에서 마이크를 이용하여 수신할 수 있다.
- [0115] 동작 662에서, 지능형 서버(520)는 추출된 제1 텍스트 데이터에 적어도 일부 기초하여, 응답 모델들 중 적어도 하나를 포함하는 리스트를 사용자에게 제공할 수 있다.
- [0116] 동작 663에서, 지능형 서버(520)는 응답 모델 리스트에서 적어도 하나의 응답 모델을 선택하는 사용자 입력을 수신할 수 있다.
- [0117] 동작 664에서, 지능형 서버(520)는 수신된 사용자 입력에 따라 적어도 하나의 응답 모델을 선택할 수 있다.
- [0118] 도 6은 일 실시 예에 따른 사용자 단말이 호출어를 등록하는 화면을 나타낸 도면이다.
- [0119] 도 6을 참조하면, 사용자 단말(510)은 디스플레이(예: 도 1의 디스플레이(140))를 통해 호출어를 등록하기 위한 UI(사용자 인터페이스: user interface)를 출력할 수 있다.
- [0120] 일 실시 예에 따르면, 710 화면에서, 사용자 단말(510)은 호출어 등록을 시작하기 위한 UI를 출력할 수 있다. 사용자 단말(510)은 호출어를 등록하기 위한 사용자 발화의 가이드 정보(711)를 상기 UI에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 오브젝트(예: 가상의 버튼)(713)를 통해 호출어 등록을 시작하기 위한 사용자 입력을 수신할 수 있다.
- [0121] 일 실시 예에 따르면, 720 화면에서, 사용자 단말(510)은, 도 5의 610 동작을 실행하는 단계에서, 호출어 등록을 위한 제1 음성 입력을 수신하기 위한 UI를 출력할 수 있다. 사용자 단말(510)은 제1 음성 입력을 수신하기 위한 제1 안내 정보(721), 및 호출어 등록 단계를 표시하기 위한 인디케이터(723)를 상기 UI에 표시할 수 있다. 인디케이터(723)는 제1 음성 입력을 수신하는 단계를 나타낼 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 제1 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(510)은 “안녕, 빅스비!”를 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 지능형 서버(예: 도 4의 지능형 서버(520))를 통해 상기 제1 음성 입력이 지정된 단어를 포함하고 있는지 판단할 수 있다.
- [0122] 일 실시 예에 따르면, 730 화면에서, 사용자 단말(510)은, 도 5의 630 동작을 실행하는 단계에서, 제1 음성 입력과 동일한 단어를 포함하는 제2 음성 입력을 수신하기 위한 UI를 출력할 수 있다. 사용자 단말(510)은 제2 음성 입력을 수신하기 위한 제2 안내 정보(731), 및 제2 음성 입력을 수신하는 단계를 나타내는 인디케이터(733)를 상기 UI에 표시할 수 있다. 제2 안내 정보(731)는, 예를 들어, 제1 음성 입력에 대응되는 텍스트 데이터(예: 안녕, 빅스비)가 포함될 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 제2 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(510)은 “안녕, 빅스비!”를 다시 수신할 수 있다.
- [0123] 일 실시 예에 따르면, 740 화면에서, 사용자 단말(510)은 도 5의 650 동작을 실행하는 단계에서, 호출어 인식 훈련을 수행하는 과정을 나타낸 UI를 출력할 수 있다. 사용자 단말(510)은 훈련 중임을 표시하는 제3 안내 정보(741), 및 훈련 단계를 나타내는 인디케이터(743)를 상기 UI에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 제1 음성 입력, 및 상기 제2 음성 입력에 기초하여 호출어를 등록할 수 있다. 예를 들어, 사

용자 단말(510)은 상기 제1 음성 입력, 및 상기 제2 음성 입력에 기초하여 호출어 인식 모델을 생성할 수 있다. 사용자 단말(510)은 생성된 호출어 인식 모델을 메모리(예: 도 4의 메모리(1311))에 저장할 수 있다.

- [0124] 일 실시 예에 따르면, 750 화면에서, 사용자 단말(510)은 호출어를 등록한 결과를 나타내는 UI를 출력할 수 있다. 사용자 단말(510)은 호출어 등록 결과를 포함하는 제3 안내 정보(751), 및 지정된 기능을 수행하기 위한 음성 입력의 예시(753)를 상기 UI에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 오브젝트(예: 가상의 버튼)(755)를 통해 호출어 등록을 완료하기 위한 사용자 입력을 수신할 수 있다.
- [0125] 이에 따라, 사용자 단말(510)이 등록된 호출어를 인식하였을 때, 음성 인식 서비스(또는, 음성 기반의 지능형 어시스턴트 서비스)를 대기 상태에서 활성화 상태로 변경될 수 있다.
- [0126] 도 7은 일 실시 예에 따른 통합 지능 시스템이 응답 모델을 설정하는 방법을 나타낸 순서도(800)이다.
- [0127] 도 7을 참조하면, 통합 지능 시스템(예: 도 4의 통합 지능 시스템(500))은 등록된 호출어에 기초하여 응답 모델을 결정할 수 있다.
- [0128] 일 실시 예에 따르면, 811 동작에서, 사용자 단말(510)은 도 5의 610 동작과 유사하게 제1 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(510)은 “김비서”를 수신할 수 있다. 일 실시 예에 따르면, 813 동작에서, 사용자 단말(510)은 상기 수신된 제1 음성 입력을 지능형 서버(520)로 송신할 수 있다.
- [0129] 일 실시 예에 따르면, 821 동작에서, 지능형 서버(520)는 도 5의 620 동작과 유사하게 제1 음성 입력에 대한 음성 인식을 수행할 수 있다. 예를 들어 지능형 서버(520)는 제1 음성 입력을 텍스트 데이터인 '김비서'로 변경할 수 있다. 일 실시 예에 따르면, 823 동작에서, 지능형 서버(520)는 인식된 결과를 사용자 단말(510)로 송신할 수 있다. 상기 인식된 결과는, 예를 들어, 제1 음성 입력에 대응되는 텍스트 데이터(예: '김비서')를 포함할 수 있다.
- [0130] 일 실시 예에 따르면, 831 동작에서, 사용자 단말(510)은 도 5의 630 동작과 유사하게 사용자에게 추가 발화를 요청할 수 있다. 예를 들어, 사용자 단말(510)은 “김비서를 3번 더 발화해주세요.”라고 출력할 수 있다. 일 실시 예에 따르면, 833 동작에서, 사용자 단말(510)은 도 5의 640 동작과 유사하게 제1 음성 입력과 동일한 제2 음성 입력을 수신할 수 있다. 또한, 사용자 단말(510)은 제1 음성 입력과 동일한 복수의 음성 입력을 수신할 수 있다. 일 실시 예에 따르면, 835 동작에서, 사용자 단말(510)은 도 5의 650 동작과 유사하게 제1 음성 입력 및 제2 음성 입력에 기초하여 음성 인식 모델을 생성할 수 있다. 다시 말해, 사용자 단말(510)은 음성 인식 모델을 훈련시킬 수 있다. 예를 들어, 사용자 단말(510)은 “김비서”에 대응되는 음성 인식 모델을 생성할 수 있다.
- [0131] 일 실시 예에 따르면, 841 동작 내지 845 동작에서, 지능형 서버(520)는 사용자 단말(510)이 음성 인식 모델을 훈련시킬 때, 응답 모델을 선택할 수 있다. 일 실시 예에 따르면, 841 동작에서, 지능형 서버(520)는 인식된 호출어에 기초하여 응답 모델을 선택할 수 있다. 예를 들어, 지능형 서버(520)는 “김비서”와 의미가 비슷한 응답 모델(예: 신중하고 예의 바른 어시스턴트 응답 모델, 발랄하고 수다스러운 여자친구 응답 모델, 따뜻하고 부드러운 남자친구 응답 모델 등)을 선택할 수 있다. 일 실시 예에 따르면, 843 동작에서, 지능형 서버(520)는 추천 모델(또는, 추천 응답 모델) 정보를 생성할 수 있다. 상기 추천 모델 정보는, 예를 들어, 추천 모델 리스트를 포함할 수 있다. 상기 리스트는 상기 선택된 적어도 하나의 응답 모델을 포함할 수 있다. 일 실시 예에 따르면, 845 동작에서, 지능형 서버(520)는 상기 생성된 추천 모델 정보를 사용자 단말(510)로 송신할 수 있다.
- [0132] 일 실시 예에 따르면, 851 동작에서, 사용자 단말(510)은 응답 모델을 선택하기 위한 사용자 입력을 수신할 수 있다. 예를 들어, 사용자 단말(510)은 추천 모델 리스트에 포함된 적어도 하나의 응답 모델 중 적어도 하나의 음성 응답(예: 신중하고 예의 바른 수행 어시스턴트 응답)을 선택하는 사용자 입력을 수신할 수 있다. 일 실시 예에 따르면, 853 동작에서, 사용자 단말(510)은 사용자 입력에 의한 선택 정보를 지능형 서버(520)로 송신할 수 있다.
- [0133] 일 실시 예에 따르면, 860 동작에서, 지능형 서버(520)는 상기 선택 정보에 대응되는 응답 모델을 설정할 수 있다. 예를 들어, 지능형 서버(520)는 선택 정보에 대응되는 응답 모델을 이용하여 응답을 생성하도록 설정될 수 있다. 이에 따라, 지능형 서버(520)는 설정된 음성 응답 모델을 이용하여 수신된 음성 입력에 대응되는 응답을 생성할 수 있다.
- [0134] 도 8은 일 실시 예에 따른 통합 지능 시스템이 사용자 단말의 초기 응답 모델을 설정하는 방법을 나타낸 도면이다.
- [0135] 도 8을 참조하면, 통합 지능 시스템(예: 도 4의 통합 지능 시스템(500))은 사용자 단말(510)의 초기 설정 시

(out of box experience)(OOBE)에 사용자 정보를 이용하여 응답 모델을 설정하기 위한 가이드 정보를 출력할 수 있다.

- [0136] 일 실시 예에 따르면, 사용자 단말(510)은 로그인(log-in)된 사용자 정보(910)를 이용하여 응답 모델을 설정하도록 가이드 정보를 출력할 수 있다. 예를 들어, 사용자 단말(510)은 사용자 정보에 기초하여 결정된 UI(user interface)(920)를 통해 응답 모델을 가이드 하는 텍스트 정보(921)를 사용자에게 제공할 수 있다. 또한, 사용자 단말(510)은 응답 모델을 가이드 하는 음성 정보를 사용자에게 제공할 수 있다. 상기 응답 모델을 설정하도록 가이드 하는 UI(920)의 색상(예: 밝은 색상), 텍스트 정보(921)의 문체(예: 친근한 문체), 상기 음성 정보의 어투(예: 친근한 어투)는 사용자 정보에 기초하여 결정될 수 있다. 일 실시 예에 따르면, 사용자 정보는 사용자의 신상 정보뿐만 아니라, 사용자 선호 정보를 포함할 수 있다. 예를 들어, 사용자 정보는 성별(예: 여성), 나이(예: 10대), 위치(예: 학교 및 집), 자주 사용하는 앱(예: 유튜브, 페이스북), 이벤트 히스토리(예: 중간고사), 및 서치 데이터(예: 브이앱)를 포함할 수 있다.
- [0137] 도 9a 및 도 9b는 일 실시 예에 따른 통합 지능 시스템이 사용자 단말의 초기 응답 모델을 설정하는 방법을 나타낸 도면들이다.
- [0138] 도 9a를 참조하면, 통합 지능 시스템(예: 도 4의 통합 지능 시스템(500))은 사용자 단말(510)의 초기 설정(out of box experience)(OOBE)시에 호출어를 설정하기 위한 사용자 입력을 수신할 수 있다.
- [0139] 일 실시 예에 따르면, 사용자 단말(510)은 호출어 입력을 위한 UI(1010)를 디스플레이에 표시할 수 있다. 상기 디스플레이에 표시된 UI(1010)의 색상(예: 밝은 색)은 사용자 정보에 기초하여 결정될 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 디스플레이에 표시된 UI(1010)을 통해 호출어를 입력하는 사용자 입력(1011)을 수신할 수 있다. 사용자 입력(1011)은, 예를 들어, 텍스트 데이터(예: 나의 아저씨)를 직접 입력할 수 있다. 일 실시 예에 따르면, 상기 입력된 텍스트 데이터는 그 자체로 호출명으로 사용될 수 있다.
- [0140] 일 실시 예에 따르면, 지능형 서버(예: 도 5의 지능형 서버(520))는 수신된 사용자 입력(1011)에 기초하여 응답 모델을 선택할 수 있다. 또한, 지능형 서버는 사용자 입력(1011)과 함께, 사용자 정보를 이용하여 응답 모델을 선택할 수 있다. 일 실시 예에 따르면, 지능형 서버는 선택된 응답 모델을 이용한 응답을 제공할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 선택된 응답 모델에 기초하여 UI(1020)를 디스플레이에 표시할 수 있다. 사용자 단말(510)은 UI(1020)를 통해 선택된 응답 모델을 안내하는 텍스트 정보(1021)를 표시할 수 있다. UI(910)의 색상(예: 어두운 색상), 텍스트 정보(921)의 문체(예: 격식 있는 문체), 상기 음성 정보의 어투(예: 격식 있는 어투)는 선택된 응답 모델에 따라 결정될 수 있다.
- [0141] 도 9b를 참조하면, 통합 지능 시스템은 사용자 단말(510)의 초기 설정 시에 응답 모델을 선택하는 사용자 입력을 수신할 수 있다.
- [0142] 일 실시 예에 따르면, 사용자 단말(510)은 호출어 입력을 위한 UI(1010)를 디스플레이에 표시할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 디스플레이에 표시된 UI(1010)을 통해 호출어를 입력하는 사용자 입력(1011)을 수신할 수 있다.
- [0143] 일 실시 예에 따르면, 지능형 서버는 수신된 사용자 입력(1011)에 기초하여 적어도 하나의 응답 모델을 선택할 수 있도록 적어도 하나의 응답 모델을 포함하는 리스트를 사용자 단말(510)로 송신할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 응답 모델을 선택하는 사용자 입력을 수신하기 위한 리스트(1031)를 디스플레이에 표시된 UI(1030)에 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 디스플레이에 표시된 리스트(1031)를 통해 응답 모델을 선택하는 사용자 입력(1031a)을 수신할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 수신된 사용자 입력(1031a)을 지능형 서버로 송신할 수 있다.
- [0144] 일 실시 예에 따르면, 지능형 서버는 수신된 사용자 입력(1031a)에 따라 응답 모델을 설정할 수 있다. 일 실시 예에 따르면, 지능형 서버는 설정된 응답 모델을 이용한 응답을 제공할 수 있다. 일 실시 예에 따르면, 사용자 단말(510)은 상기 선택된 응답 모델에 기초하여 UI(1020)를 디스플레이에 표시할 수 있다.
- [0145] 도 10a 및 도 10b는 일 실시 예에 따른 응답 모델이 통합 지능 시스템에 설정되는 방법을 나타낸 도면들이다.
- [0146] 도 10a를 참조하면, 통합 지능 시스템(예: 도 4의 통합 지능 시스템(500))은 호출어에 기초하여 선택된 응답 모델(529b_N)이 설정될 수 있다.
- [0147] 일 실시 예에 따르면, 응답 모델 선택 모듈(1110)은 지능형 서버(예: 도 4의 지능형 서버(520))의 복수의 구성 중 적어도 일부로 구현될 수 있다. 다시 말해, 응답 모델 선택 모듈(1110)의 기능은 지능형 서버의 복수의 구성

중 적어도 일부를 통해 구현될 수 있다.

- [0148] 일 실시 예에 따르면, 응답 모델 선택 모듈(1110)은 호출어에 기초하여 응답 모델(529b_N)을 선택할 수 있다. 예를 들어, 응답 모델 선택 모듈(1110)은 호출어에 기초하여 복수의 응답 모델(529b_1 내지 529b_M) 중 적어도 하나의 응답 모델(529b_N)을 선택할 수 있다. 복수의 응답 모델(529b_1 내지 529b_M)은, 예를 들어, 응답 모델 데이터베이스(529b)에 저장될 수 있다. 호출어에 기초하여 응답 모델(529b_N)을 선택하는 방법은 도 11에서 자세히 설명하도록 하겠다.
- [0149] 일 실시 예에 따르면, 응답 모델 선택 모듈(1110)은 선택된 응답 모델(529b_N)을 통합 지능 시스템(500)에 설정할 수 있다. 이에 따라, 통합 지능 시스템(500)은 설정된 응답 모델(529b_N)을 이용하여 수신된 사용자 입력에 대응되는 응답을 생성할 수 있다.
- [0150] 도 10b를 참조하면, 선택된 응답 모델(529b_N)은 통합 지능 시스템(500)에 포함된 각 구성에 대한 모델을 포함할 수 있다.
- [0151] 일 실시 예에 따르면, 응답 모델(529b_N)은 지능형 서버에 포함된 각 구성에 대한 모델을 포함할 수 있다. 예를 들어, 응답 모델(529_M)은 플래너 모델(529b_N1), 자연어 응답 생성 모델(529b_N3), 영상 응답 생성 모델(529b_N5), 텍스트 음성 변환(TTS) 모델(529b_N7), 또는 대화 관리 모델(529b_N9) 중 적어도 하나 이상을 포함할 수 있다.
- [0152] 일 실시 예에 따르면, 플래너 모델(529b_N1)은 지능형 서버의 플래너 모듈(523)에서 플랜을 생성할 때 이용될 수 있다. 상기 플랜은, 예를 들어, 수신된 음성 입력의 의도에 기초하여 생성될 수 있다. 상기 플랜은 사용자 단말(100)에서 실행되는 복수의 동작을 포함할 수 있다. 일 실시 예에 따르면, 플래너 모듈(523)은 플래너 모델(529b_N1)에 따라 플랜을 생성할 수 있다. 예를 들어, 플래너 모듈(523)은 플래너 모델(529b_N1)에 따라 지정된 앱을 실행하거나, 지정된 기능을 수행하는 플랜을 생성할 수 있다. 다시 말해, 플래너 모듈(523)은 선택된 응답 모델(529b_N)의 플래너 모듈(529b_N1)에 따라 동일한 의도에 대해 상이한 앱을 실행하거나, 상이한 기능을 수행하는 플랜을 생성할 수 있다. 일 실시 예에 따르면, 플래너 모듈(529b_N1)은 지정된 앱, 또는 기능에 대한 정보를 포함할 수 있다.
- [0153] 일 실시 예에 따르면, 자연어 응답 생성 모델(529b_N3)은 지능형 서버(520)의 자연어 응답 생성 모듈(524)에서 자연어 정보를 포함하는 응답을 생성할 때 이용될 수 있다. 상기 응답은, 예를 들어, 수신된 음성 입력의 의도에 기초하여 생성될 수 있다. 일 실시 예에 따르면, 자연어 응답 생성 모듈(524)은 자연어 응답 생성 모델(529b_N3)에 따라 지정된 형식으로 응답을 생성할 수 있다. 예를 들어, 자연어 응답 생성 모듈(524)은 자연어 응답 생성 모델(529b_N3)에 따라 화면에 표시될 심볼(예: 이모티콘), 또는 자연어 응답(텍스트)을 포함하는 응답을 생성할 수 있다. 다시 말해, 자연어 응답 생성 모듈(524)은 선택된 응답 모델(529b_N)의 자연어 응답 생성 모델(529b_N3)에 따라 동일한 의도에 대해 상이한 심볼 및/또는 텍스트를 포함하는 자연어 응답을 생성할 수 있다. 일 실시 예에 따르면, 자연어 응답 생성 모델(529b_N3)은 특정 응답을 생성하는 규칙을 포함할 수 있다. 상기 규칙은, 예를 들어, 화면에 포함될 자연어 응답(예: 이미지 및/또는 텍스트)을 포함할 수 있다. 상기 화면에 포함될 자연어 응답은 이모티콘을 나타내는 심볼 표현 규칙을 포함할 수 있다. 일 실시 예에 따르면, 규칙은 사용자가 사용하는 모드에 따라 설정될 수 있다. 상기 사용자가 사용하는 모드는, 예를 들어, 아이-프리 모드(eyes-free mode), 핸드-프리 모드(hands-free mode), 및 스크린 모드(screen mode)를 포함할 수 있다. 일 실시 예에 따르면, 복수의 규칙은 하나의 응답(또는, 응답 ID(identification))에 대응될 수 있다.

표 1

NLG ID	NLG ver.	Comment #10-1	dialogue #10-1-1 (Display)	dialogue #10-1-1 (Spoken)	dialogue #10-1-2 (Display)	dialogue #10-1-2 (Spoken)
Starbucks_10_CreateOrder_PaymentOption_Exist_No	0.9	[보통이] 수다, 반존칭, 추천기능	뭘로 계산할까? 지난해에 계산한 걸로?	뭘로 계산할까? 지난해에 계산한 걸로? %smile	이제 계산해볼까? 이걸로?	이제 계산해볼까? 이걸로? %question
Starbucks_11_CreateOrder_DTStore_Exist_Yes	0.8	[보통이] 수다, 반존칭, 이모티콘	#MenuName로 가는거야? 그럼 결제버튼 눌러줘	#MenuName로 가는거야? %wow 그럼 결제버튼 눌러줘 %wink	다했음 결제버튼 누르기	다했음 결제버튼 누르기 %wink

- [0154]
- [0155] 표 1과 같이 자연어 생성(NLG) 모듈(227)은 적어도 하나의 응답 ID를 가질 수 있다. 예를 들어, 스타벅스 10 및 스타벅스 11이라는 응답 ID를 가질 수 있다. 각각의 응답 ID에는 대응하는 자연어 생성 버전, 코멘트(comment)

테마 및 복수의 대화(dialogue) 표시(display) 및/또는 발화(spoken)에 관련된 규칙들이 지정될 수 있다. 예를 들어, 코멘트 테마가 “보통이” 인 경우, 수다, 반존칭, 추천기능, 및/또는 이모티콘과 관련된 대화 표시 또는 발화 규칙들을 지정할 수 있다.

표 2

Comment #10-2	dialogue #10-2-1 (Display)	dialogue #10-2-1 (Spoken)	dialogue #10-2-2 (Display)	dialogue #10-2-2 (Spoken)
[나의 아저씨] 단호하고 짧은 정갈한 표현.	원하시는 결제 수단을 선택해주세요.	원하시는 결제 수단을 선택해주세요.	어떤 방법으로 결제할지 선택해주세요.	어떤 방법으로 결제할지 선택해주세요.
[나의 아저씨] 단호하고 짧은 정갈한 표현.	원하시는 결제 버튼을 눌러주세요.	결제 버튼을 눌러주세요.	주문을 마치려면 결제 버튼을 눌러주세요.	주문을 마치려면 결제 버튼을 눌러주세요.

[0156]

[0157]

표 2와 같이 각각의 응답 ID의 코멘트 테마를 변화시키는 경우, 복수의 대화(dialogue) 표시(display) 및/또는 발화(spoken)에 관련된 다른 규칙들이 지정될 수 있다. 예를 들어, 코멘트 테마가 “나의 아저씨” 인 경우, 동일한 상황에 대하여 단호하고 짧은 표현 및/또는 정갈한 표현으로 대화 표시 또는 발화 규칙들을 지정할 수 있다. 일 실시 예에 따르면, 영상 응답 생성 모듈(529b_N5)은 지능형 서버(520)의 영상 응답 생성 모듈(525)에서 영상 정보를 포함하는 응답을 생성할 때 사용될 수 있다. 상기 응답은, 예를 들어, 수신된 음성 입력의 의도에 기초하여 생성될 수 있다. 일 실시 예에 따르면, 영상 응답 생성 모듈(525)은 영상 응답 생성 모듈(529b_N5)에 따라 지정된 형식으로 응답을 생성할 수 있다. 예를 들어, 영상 응답 생성 모듈(525)은 영상 응답 생성 모듈(529b_N5)에 따라 콘텐츠(content)를 선택하거나, 지정된 창의 형태 및 색상의 응답을 생성할 수 있다. 다시 말해, 영상 응답 생성 모듈(525)은 선택된 응답 모듈(529b_N)의 영상 응답 생성 모듈(529b_N5)에 따라 동일한 의도에 대해 상이한 콘텐츠를 선택하거나, 상이한 창의 형태 및 색상의 응답을 생성할 수 있다. 일 실시 예에 따르면, 영상 응답 생성 모듈(529b_N5)은 콘텐츠 선택 규칙, 또는 창의 기본 형태 및 색상에 대한 정보를 포함할 수 있다.

[0158]

영상 응답 생성 모듈(529_N5)에 포함된 창의 형태와 색상은 사용자 명령에 대한 응답을 생성할 때 선택 또는 사용될 수 있다. 뿐만 아니라, 영상 응답 생성 모듈(529_N5)에 포함된 창의 기본 형태 및/또는 색상은 사용자가 호출어를 입력한 이후 사용자 명령이 입력되기 전 대기 상태를 나타내는 화면을 표시할 때, 대기 화면을 표시하기 위한 정보로 사용될 수 있다.

[0159]

일 실시 예에 따르면, 텍스트 음성 변환 모듈(529b_N7)은 지능형 서버(520)의 텍스트 음성 변환 모듈(526)에서 자연어 응답 생성 모듈(524)에서 생성된 자연어 정보를 음성 신호로 변환할 때 사용될 수 있다. 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 텍스트 음성 변환 모듈(529b_N7)에 따라 자연어 정보를 지정된 형식의 음성 신호로 변환할 수 있다. 예를 들어, 텍스트 음성 변환 모듈(526)은 텍스트 음성 변환 모듈(529b_N7)에 포함된 합성 단위(예: 음소(phoneme), 다이폰(diphone), 트라이폰(triphone) 등) 별 음파신호(waveform), 및 파라미터(예: 피치(pitch), 스펙트럼(spectrum), 대역 노이즈(band noise), 발화 속도, 억양 변화폭 등)에 따라 자연어 정보를 음성 신호로 변환할 수 있다. 다시 말해, 텍스트 음성 변환 모듈(526)은 동일한 자연어 정보를 선택된 텍스트 음성 변환 모듈(529b_N7)에 포함된 합성 단위 별 음파신호 및 파라미터에 따라 상이한 음성 신호로 변경할 수 있다. 일 실시 예에 따르면, 텍스트 음성 변환 모듈(529b_N7)은 복수의 합성 단위 별 음파신호 및 파라미터에 대한 정보를 포함할 수 있다. 예를 들어, 텍스트 음성 변환 모듈(529b_N7)은 합성될 목소리의 특성에 따라 음성 DB에 포함된 복수의 합성 단위 별 음파신호 및 파라미터 세트(예: 신호를 합성하기 위한 파라미터 세트)를 이용할 수 있다.

[0160]

일 실시 예에 따르면, 대화 관리 모듈(529b_N9)은 지능형 서버(520)의 대화 관리 모듈(527)이 대화의 흐름을 관리할 때 사용할 수 있다. 일 실시 예에 따르면, 대화 관리 모듈(527)은 대화 관리 모듈(529b_N9)에 따라 대화의 흐름을 관리할 수 있다. 예를 들어, 대화 관리 모듈(527)은 대화 관리 모듈(529b_N9)의 지정된 컨텍스트에 따라 응답을 결정할 수 있다. 다시 말해, 대화 관리 모듈(527)은 선택된 응답 모듈(529b_N)의 대화 관리 모듈

(529b_N9)에 따라 동일한 사용자 입력에 대응되는 상이한 응답을 결정할 수 있다. 일 실시 예에 따르면, 대화 관리 모델(529b_N9)은 지정된 컨텍스트에 대한 정보를 포함할 수 있다.

[0161] 일 실시 예에서, 응답 모델(529b_N)에 포함된 플래너 모델(529b_N1), 자연어 응답 생성 모델(529b_N3), 영상 응답 생성 모델(529b_N5), TTS 모델(529b_N7), 및/또는 대화 관리 모델(529b_N9)을 이용하여 응답 테마를 결정할 수 있다. 응답 모델(529b_N)에 포함된 모델들 중 일부의 모델만이 테마 결정에 사용될 수 있다. 예를 들어, 다른 모델은 그대로 사용하고 TTS 모델(529b_N7)만 테마에 따라 다르게 적용할 수 있다. 이에 따라, 통합 지능 시스템(500)은 선택된 응답 모델(529b_N)에 포함된 모델을 이용하여 사용자 입력에 대응되는 응답을 생성할 수 있다.

[0162] 도 11은 일 실시 예에 따른 지능형 서버가 응답 모델을 결정하는 방법을 나타낸 흐름도(1200)이다.

[0163] 도 11을 참조하면, 지능형 서버(예: 도 4의 지능형 서버(520))는 호출어에 기초하여 응답 모델을 결정할 수 있다.

[0164] 일 실시 예에 따르면, 1210 동작에서, 지능형 서버(예: 도 4의 자동 음성 인식 모듈(521))은 호출어를 포함하는 음성 입력을 텍스트 데이터로 변경할 수 있다. 상기 음성 입력은, 예를 들어, 호출어를 등록하기 위한 사용자 입력일 수 있다. 일 실시 예에 따르면, 지능형 서버(520)은 상기 변환된 텍스트 데이터를 이용하여 호출어를 등록할 수 있다.

[0165] 일 실시 예에서는, 도 9b의 사용자 입력(1011)과 같이, 사용자는 호출어를 텍스트로 입력할 수 있다. 호출어를 텍스트로 입력하는 경우, 별도의 변경 없이 사용자 입력이 완료될 수 있으며, 이 경우 1210 동작은 생략될 수 있다.

[0166] 일 실시 예에 따르면, 1220 동작에서, 지능형 서버(예: 도 4의 호출어 정제화 모듈(528))은 등록된 호출어를 정제하여 사용자 단말(510)의 호출명을 생성할 수 있다. 상기 호출명은, 예를 들어, 사용자 입력에 대응되는 응답에 이용될 수 있다. 예를 들어, 입력된 호출어가 "안녕 보롱이"라고 입력된 경우, 호출어 정제화 모듈은 동사에 판단되는 "안녕"은 제거하고 "보롱이"을 호출명으로 설정할 수 있다.

[0167] 일 실시 예에 따르면, 1230 동작에서, 지능형 서버는 호출명의 의미를 분석할 수 있다. 일 실시 예에 따르면, 1230a 동작에서, 지능형 서버는 호출명의 형태소를 분석할 수 있다. 다시 말해, 지능형 서버는 언어적 특징을 이용하여 호출명을 분석할 수 있다. 일 실시 예에 따르면, 1230b 동작에서, 지능형 서버는 호출명의 형태소를 이용하여 의미 구분 태깅(semantic category tagging)을 수행할 수 있다. 상기 의미 구분 태깅은, 예를 들어, 통계적 방법(예: Ciaramita, et al., "Supersense tagging of unknown nouns in WordNet," EMNLP 2003), 규칙 기반 방법(예: Curran, et al., "Supersense tagging of unknown nouns using semantic similarity," ACL 2005) 등을 포함하는 다양한 알고리즘을 이용하여 수행될 수 있다.

[0168] 일 실시 예에 따르면, 1240 동작에서, 지능형 서버는 상기 분석된 호출명의 의미와 응답 모델의 테마 사이의 의미상 거리(semantic distance)를 계산할 수 있다. 예를 들어, 지능형 서버는 중복된 태그의 개수나, 의미 네트워크(semantic network)의 거리 등을 이용하여 상기 분석된 호출명의 의미와 응답 모델의 테마 사이의 의미상 거리를 계산할 수 있다. 상기 응답 모델의 테마는, 예를 들어, 의미 구분(semantic categories) 태그를 포함할 수 있다. 이하의 표 1은 응답 모델의 테마에 설정된 의미 구분 태그(예: #의미)를 나타낸 것이다.

표 3

Theme_ID	Theme ver.	Semantic_category tags
Theme_ID_01_Borong	1.1	#chatty, #cute, #girl, #lovely, #friendly, #girl
Theme_ID_02_MyUncle	1.0	#sober, #reliable, #confident, #reticent, #attentive, #gentleman
...

[0170] 일 실시 예에 따르면, 1250 동작에서, 지능형 서버는 호출명과의 의미상 거리가 가까운 응답 모델을 선택할 수 있다. 일 실시 예에 따르면, 지능형 서버는 복수의 응답 모델이 포함된 리스트를 사용자에게 제공할 수 있다. 예를 들어, 지능형 서버는 의미상 거리가 지정된 값 이하인 응답 모델이 포함된 리스트를 사용자에게 제공할 수 있다. 다른 예를 들어, 지능형 서버는 의미상 거리가 가까운 순서를 기준으로 지정된 순위까지의 응답 모델이 포함된 리스트를 사용자에게 제공할 수 있다. 일 실시 예에 따르면, 지능형 서버는 리스트에 포함된 복수의 응

답 모델 중 적어도 하나를 선택하는 사용자 입력을 수신할 수 있다. 일 실시 예에 따르면, 지능형 서버는 상기 사용자 입력에 따라 응답 모델을 선택할 수 있다.

[0171] 이에 따라, 지능형 서버는 상기 선택된 응답 모델을 통합 지능 시스템(500)에 설정할 수 있다.

[0172] 예를 들어, 지능형 서버는 의미상 거리가 지정된 값 이상인 응답 모델을 포함하는 리스트를 사용자에게 제공할 수 있다. 다른 예를 들어, 지능형 서버는 지정된 위의 응답 모델을 포함하는 리스트를 사용자에게 제공할 수 있다. 지능형 서버는 상기 리스트를 통해 적어도 하나의 응답 모델을 선택하는 사용자 입력을 수신할 수 있다. 지능형 서버는 수신된 사용자 입력에 대응되는 응답 모델을 선택할 수 있다.

[0173] 도 12는 일 실시 예에 따른 통합 지능 시스템(500)이 음성 입력에 대한 응답을 제공하는 방법을 나타낸 흐름도(1300)이다.

[0174] 도 12를 참조하면, 통합 지능 시스템(500)은 호출어에 기초하여 선택된 응답 모델이 설정될 수 있다. 통합 지능 시스템(500)의 사용자 단말(510)은 상기 호출어가 포함된 음성 입력을 수신하면, 사용자 단말(510)에서 제공하는 음성 인식 서비스 또는 음성 기반의 지능형 어시스턴트 서비스를 활성화시킬 수 있다.

[0175] 일 실시 예에 따르면, 1311 동작에서, 사용자 단말(510)의 서브 프로세서(513)는 호출어를 포함하는 음성 입력을 수신할 수 있다. 예를 들어, 사용자 단말(510)은 “김비서”를 포함하는 음성 입력을 수신할 수 있다.

[0176] 일 실시 예에 따르면, 1313 동작에서, 서브 프로세서(513)는 음성 입력에 포함된 호출어를 인식할 수 있다. 일 실시 예에 따르면, 서브 프로세서(513)는 사용자 단말(510)의 메인 프로세서(515)로 활성화 요청을 전달할 수 있다. 또한, 서브 프로세서(513)는 상기 활성화 요청과 함께 호출어에 대한 정보를 전달할 수 있다.

[0177] 일 실시 예에 따르면, 1321 동작에서, 메인 프로세서(515)는 호출어가 포함된 음성 입력을 재인식할 수 있다. 메인 프로세서(515)의 재인식 동작은 선택적으로 적용될 수 있다. 메인 프로세서(515)는 서브 프로세서(513)와 비교하여 많은 계산을 할 수 있으므로, 정확하게 호출어를 인식할 수 있다. 이에 따라, 사용자 단말(510)은 더욱 정확하게 사용자의 활성화 명령을 인식할 수 있다. 일 실시 예에 따르면, 1323 동작에서, 메인 프로세서(515)는 호출어 정보를 지능형 서버(520)로 송신할 수 있다.

[0178] 일 실시 예에 따르면, 1331 동작에서, 지능형 서버(520)는 상기 호출어에 대한 정보를 사용자 단말(510)로부터 수신하고, 상기 수신된 호출어에 대응되는 응답 모델을 설정할 수 있다.

[0179] 일 실시 예에 따르면, 1331 동작은 사용자 단말(510)에서 전송되는 호출어 정보를 사용하지 않고, 이전의 호출어를 설정하는 동작 중에 결정된 응답 모델을 사용하는 것으로 대체될 수 있다. 이 경우, 사용자 단말(510)에서 전송되는 호출어 정보가 필요하지 않으므로 1323 동작은 생략될 수 있다. 일 실시 예에 따르면, 1341 동작에서, 메인 프로세서(515)는 지정된 기능을 수행하기 위한 명령어가 포함된 음성 입력을 수신하기 위해 대기할 수 있다. 일 실시 예에 따르면, 1343 동작에서, 메인 프로세서(515)는 명령어가 포함된 사용자 입력을 수신할 수 있다. 예를 들어, 메인 프로세서(515)는 “오늘 서울 날씨 알려줘!”라는 음성 입력을 수신할 수 있다. 일 실시 예에 따르면, 1345 동작에서, 메인 프로세서(515)는 상기 수신된 음성 입력을 지능형 서버(520)로 송신할 수 있다.

[0180] 일 실시 예에 따르면, 1351 동작에서, 지능형 서버(520)는 사용자 단말(510)로부터 지정된 기능을 수행하기 위한 음성 입력을 인식하고, 인식된 결과를 사용자 단말(510)로 송신할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 상기 음성 입력을 텍스트 데이터로 변환할 수 있다. 예를 들어, 지능형 서버(520)는 “오늘 서울 날씨 알려줘!”를 텍스트 데이터로 변환할 수 있다. 사용자 단말(510)은, 상기 인식된 텍스트 데이터를 사용자에게 제공할 수 있다.

[0181] 일 실시 예에 따르면, 1353 동작에서, 지능형 서버(520)는 설정된 응답 모델을 이용하여 수신된 음성 입력에 대응되는 응답을 생성할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 변환된 텍스트 데이터에 기초하여 의도 및 파라미터를 결정할 수 있다. 예를 들어, 지능형 서버(520)는 날씨 정보를 획득하기 위한 의도(예: intent = _WEATHER_MESSAGE_), 및 날짜(예 param.weather.date = “2018.2.28”) 또는 위치(예: param.weather.location = “seoul”)를 포함하는 파라미터를 결정할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 상기 결정된 의도 및 파라미터에 기초하여 플랜을 생성할 수 있다. 예를 들어, 지능형 서버(520)는 설정된 응답 모델을 이용하여 상기 플랜을 생성할 수 있다. 지능형 서버(520)는 상기 생성된 플랜에 따라 결과를 생성할 수 있다. 일 실시 예에 따르면, 지능형 서버(520)는 설정된 응답 모델을 이용하여 상기 생성된 결과를 포함하는 응답을 생성할 수 있다. 예를 들어, 지능형 서버(520)는 “김비서가 날씨를 알려드리겠습니다. 오

를 날씨는 오후 늦게 비가 내리겠습니다.” 라는 자연어 응답을 생성하고, 생성된 자연어 응답을 도 4의 텍스트 음성 변환 모듈(526)을 통해 음성 신호(예: 차분한 남성 목소리)로 변환하여 응답을 생성할 수 있다. 일 실시 예에 따르면, 1355 동작에서, 지능형 서버(520) 상기 생성된 응답을 사용자 단말(510)로 송신할 수 있다.

[0182] 일 실시 예에 따르면, 1361 동작에서, 사용자 단말(510)의 메인 프로세서(515)는 수신된 응답을 출력할 수 있다.

[0183] 도 13은 일 실시 예에 따른 지능형 서버의 텍스트 음성 변환 모듈의 동작을 나타낸 도면이다.

[0184] 도 13을 참조하면, 지능형 서버(예: 도 4의 지능형 서버(520))의 텍스트 음성 변환 모듈(526)은 다양한 알고리즘을 이용하여 음성 신호를 생성할 수 있다. 텍스트 음성 변환 모듈(526)은 유닛 선택 합성 알고리즘(unit selection synthesis)을 이용하여 음성 신호를 생성할 수 있다.

[0185] 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 자연어 처리부(526_1) 및 디지털 신호 처리부(526_3)을 포함할 수 있다. 일 실시 예에 따르면, 자연어 처리부(526_1)는 텍스트 데이터에 기초하여 합성에 필요한 유닛 및 속성을 결정할 수 있다. 일 실시 예에 따르면, 디지털 신호 처리부(526_3)는 상기 결정된 속성에 기초하여 음성 신호를 생성할 수 있다. 디지털 신호 처리부(526_3)는 디지털 신호 생성부(526_3a) 및 음성 데이터베이스(DB)(526_3b)를 포함할 수 있다. 일 실시 예에 따르면, 디지털 신호 생성부(526_3a)는 음성 신호 합성에 필요한 파라미터를 변경할 수 있다. 예를 들어, 신호 생성부(526_3a)는 피치(pitch), 발화 속도(예: 음성 길이), 및 음색(예: 스펙트럼 정보) 중 적어도 하나를 변경할 수 있다. 일 실시 예에 따르면, 음성 데이터베이스(DB)(526_3b)는 복수의 화자로부터 수집된 음성 데이터베이스를 저장할 수 있다. 이하의 표 2는 음성 데이터베이스(526_3b)에 저장된 음성 정보를 나타낸 것이다.

표 4

[0186]

Theme ID	Theme ver.	TTS parameter sets				
		DB	pitch bias	Intonation contrast	speed	...
Theme_ID_01_Borong	1.1	Suyoun_v1.0	-5	+5	+5	...
Theme_ID_02_MyUncle	1.0	Jinsoo_v2.0	+10	0	0	...
...

[0187] 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 선택된 응답 모델에 따라 음성 신호 합성에 필요한 파라미터를 변경할 수 있다. 예를 들어, 텍스트 음성 변환 모듈(526)의 디지털 신호 처리부(526_3)는 TD-PSOLA (time domain pitch synchronous overlap and add) 동작을 수행할 때, 음성 신호 합성에 필요한 파라미터를 변경할 수 있다. 일 실시 예에 따르면, 텍스트 음성 변환 모듈(526)은 상기 변경된 파라미터에 따라 음성 신호를 생성할 수 있다. 예를 들어, 디지털 신호 처리부(526_3)는 변경된 파라미터에 따라 음성 데이터베이스(526_3b)에 저장된 음성 정보를 이용하여 음성 신호를 생성할 수 있다.

[0188] 도 14a 및 14b는 일 실시 예에 따른 통합 지능 시스템이 선택된 응답 모델에 따른 서비스를 제공하는 것을 나타낸 도면들이다.

[0189] 도 14a 및 14b를 참조하면, 통합 지능 시스템(예: 도 4의 통합 지능 시스템(500))은 선택된 음성 모델에 따라 수신된 음성 입력에 대응되는 응답을 제공할 수 있다.

[0190] 일 실시 예에 따르면, 통합 지능 시스템(500)은 호출어에 기초하여 지정된 응답 모델로 설정될 수 있다. 일 실시 예에 따르면, 통합 지능 시스템(500)은 복수의 응답 모델에 대한 정보를 저장할 수 있다. 일 실시 예에 따르면, 응답 모델의 정보는 응답을 생성하는데 필요한 정보를 포함할 수 있다. 예를 들어, 응답 모델의 정보는 이모티콘의 사용 여부, 문장 길이, 존칭 여부, 표현 스타일, 유행어 사용여부, 텍스트 음성 변환 특징, 추천 기능, 및 특화된 기능 등 응답 메시지 생성 시 적용되는 규칙 등에 대한 정보를 포함할 수 있다. 이하의 표 3에서 “나의 아저씨” 인 호출어에 기초하여 선택된 정중한 응답 모델(예: 어시스턴트 응답 모델)과 “보롱이” 인 호출어에 기초하여 선택된 캐릭터 응답 모델(예: 귀여운 응답 모델)에 대한 메시지 생성 정보를 나타낸 것이다.

표 5

[0191]

	정중한 응답 모델	캐릭터 응답 모델
이모티콘 사용 여부	X	0

문장 길이	짧고 간결함	길고 짧음에 구애 받지 않음
존칭 사용 여부	0	X
표현 스타일	격식을 차린 표현 사용	위트 있는 표현 사용
유행어 사용 여부	X	0 (신조어 적극 반영)
TTS 특징	남성의 목소리	어린이 목소리
추천 기능	생산성 중심의 앱 기능 추천	멀티미디어 콘텐츠 추천
특화 기능	“심심해!” : 영화, 공연 소식 제 공 “노래해봐!” : 자주 듣는 노래 제 공	“심심해!” : 만화, 게임 추천 “노래해봐!” : 지정된 장르의 노래 제 공

- [0192] 일 실시 예에 따르면, 텍스트 음성 변환 특징 정보는 음성 신호 합성 시 필요한 파라미터 정보를 포함할 수 있다. 예를 들어, 텍스트 음성 변환 특징 정보는 재생 속도(예: speed: +5.), 피치(pitch)(예: pitch bias: -5), 스펙트럼 조절 등에 대한 정보를 포함할 수 있다.
- [0193] 일 실시 예에 따르면, 통합 지능 시스템(500)은 호출어에 기초하여 선택된 응답 모델을 이용하여 생성된 응답을 제공할 수 있다.
- [0194] 일 실시 예에 따르면, 사용자 단말(510)은 선택된 응답 모델의 정보를 이용하여 생성된 응답을 디스플레이에 표시할 수 있다. 예를 들어, 사용자 단말(510)은 “나의 아저씨”라는 호출어에 기초하여 결정된 정중한 테마의 응답을 사용자에게 제공할 수 있다. 사용자 단말(510)은 어두운 색상의 배경화면이 포함된 UI(1410)을 통해 문장의 길이가 간결하고 격식을 차린 표현(예: 원하는 메뉴가 맞다면 결제 버튼을 누르세요!)을 포함하는 응답(1411)을 사용자에게 제공할 수 있다. 다른 예를 들어, 사용자 단말(510)은 “보롱이”라는 호출어에 기초하여 결정된 캐릭터 테마의 응답을 사용자에게 제공할 수 있다. 사용자 단말(510)은 밝은 색상의 배경화면이 포함된 UI(1420)를 통해 문장의 길이가 길고, 위트 있는 표현(예: 이대로 고? 최근 주문 음료에 기억해둘게요~)을 포함한 응답(1421)을 사용자에게 제공할 수 있다.
- [0195] 일 실시 예에 따르면, 통합 지능 시스템(500)은 선택된 응답 모델에 기초하여 지정된 서비스를 제공할 수 있다.
- [0196] 일 실시 예에 따르면, 통합 지능 시스템(500)은 선택된 응답 모델에 기초하여 지정된 음성 입력에 대한 결과(또는, 응답)를 제공할 수 있다. 예를 들어, 통합 지능 시스템(500)은 “나의 아저씨”가 호출어로 설정되어 있는 경우, “심심해!”라는 사용자 입력에 대해 '영화 및 공연 정보'를 포함하는 결과를 제공할 수 있다. 사용자 단말(510)은 어두운 색상의 배경화면이 포함된 UI(1430)를 통해 '영화 및 공연 정보'를 포함하는 결과(1431)를 제공할 수 있다. 다른 예를 들어, 통합 지능 시스템(500)은 “보롱이”가 호출어로 설정되어 있는 경우, “심심해!”라는 사용자 입력에 대해 '만화, 게임 정보'를 포함하는 결과를 제공할 수 있다. 사용자 단말(510)은 밝은 색상의 배경화면이 포함된 UI(1440)을 통해 '만화, 게임 정보'를 포함하는 결과(1441)를 제공할 수 있다.
- [0197] 본 발명의 도 1 내지 도 13에서 설명한 통합 지능 시스템(500)은 사용자가 설정 가능한 호출어에 기초하여 응답 모델을 선택하고, 선택된 응답 모델을 이용하여 음성 응답을 생성함으로써, 개인화된 음성 어시스턴트 서비스를 제공할 수 있다. 이에 따라, 인간과 대화하는 것과 같은 사용자 경험성, 및 사용자 단말에 대한 사용자의 소유 감뿐만 아니라, 음성 어시스턴트 서비스의 사용성을 증가시킬 수 있다.
- [0198] 도 15는 다양한 실시 예들에 따른 네트워크 환경(1500) 내의 전자 장치(1501)의 블록도이다. 본 문서에 개시된 다양한 실시 예들에 따른 전자 장치는 다양한 형태의 장치가 될 수 있다. 전자 장치는, 예를 들면, 휴대용 통신 장치(예: 스마트폰), 컴퓨터 장치(예: PDA(personal digital assistant), 태블릿 PC(tablet PC), 랩탑 PC, 데스크톱 PC, 워크스테이션, 또는 서버), 휴대용 멀티미디어 장치(예: 전자 책 리더기 또는 MP3 플레이어), 휴대용 의료 기기(예: 심박, 혈당, 혈압, 또는 체온 측정기), 카메라, 또는 웨어러블 장치 중 적어도 하나를 포함할 수 있다. 웨어러블 장치는 액세서리 형(예: 시계, 반지, 팔찌, 발찌, 목걸이, 안경, 콘택트 렌즈, 또는 머리 착용 형 장치(head-mounted-device(HMD))), 직물 또는 의류 일체형(예: 전자 의복), 신체 부착 형(예: 스킨 패드 또는 문신), 또는 생체 이식 형 회로 중 적어도 하나를 포함할 수 있다. 어떤 실시 예들에서, 전자 장치는, 예를 들면, 텔레비전, DVD(digital video disk) 플레이어, 오디오 장치, 오디오 액세서리 장치(예: 스피커, 헤드폰, 또는 헤드 셋), 냉장고, 에어컨, 청소기, 오븐, 전자레인지, 세탁기, 공기 청정기, 셋톱 박스, 홈 오토메이션 컨트롤 패널, 보안 컨트롤 패널, 게임 콘솔, 전자 사전, 전자 키, 캠코더, 또는 전자 액자 중 적어도 하나를 포함할 수 있다.

- [0199] 다른 실시 예에서, 전자 장치는 네비게이션 장치, 위성 항법 시스템(GNSS(global navigation satellite system)), EDR(event data recorder)(예: 차량/선박/비행기 용 블랙박스(black box)), 자동차 인포테인먼트 장치(예: 차량용 헤드-업 디스플레이), 산업용 또는 가정용 로봇, 드론(drone), ATM(automated teller machine), POS(point of sales) 기기, 계측 기기(예: 수도, 전기, 또는 가스 계측 기기), 또는 사물 인터넷 장치(예: 전구, 스프링클러 장치, 화재 경보기, 온도 조절기, 또는 가로등) 중 적어도 하나를 포함할 수 있다. 본 문서의 실시 예에 따른 전자 장치는 전술한 기기들에 한정되지 않으며, 또한, 예를 들면, 개인의 생체 정보(예: 심박 또는 혈당)의 측정 기능이 구비된 스마트폰의 경우처럼, 복수의 장치들의 기능들을 복합적으로 제공할 수 있다. 본 문서에서, 사용자라는 용어는 전자 장치를 사용하는 사람 또는 전자 장치를 사용하는 장치(예: 인공지능 전자 장치)를 지칭할 수 있다.
- [0200] 도 1을 참조하여, 네트워크 환경(1500)에서 전자 장치(1501)(예: 도 1의 사용자 단말(100), 및 도 4의 사용자 단말(510))은 근거리 무선 통신(1598)을 통하여 전자 장치(1502)와 통신하거나, 또는 네트워크(1599)를 통하여 전자 장치(1504) 또는 서버(1508)와 통신할 수 있다. 일 실시 예에 따르면, 전자 장치(1501)는 서버(1508)을 통하여 전자 장치(1504)와 통신할 수 있다.
- [0201] 일 실시 예에 따르면, 전자 장치(1501)는 버스(1510), 프로세서(1520)(예: 도 1의 프로세서(160), 도 4의 서브 프로세서(513), 및 메인 프로세서(515)), 메모리(1530)(예: 도 1의 메모리(150), 및 도 4의 메모리(511)), 입력 장치(1550)(예: 마이크 또는 마우스), 표시 장치(1560), 오디오 모듈(1570), 센서 모듈(1576), 인터페이스(1577), 햅틱 모듈(1579), 카메라 모듈(1580), 전력 관리 모듈(1588), 및 배터리(1589), 통신 모듈(1590), 및 가입자 식별 모듈(1596)을 포함할 수 있다. 어떤 실시 예에서는, 전자 장치(1501)는 구성요소들 중 적어도 하나(예: 표시 장치(1560) 또는 카메라 모듈(1580))를 생략하거나 다른 구성요소를 추가적으로 구비할 수 있다.
- [0202] 버스(1510)는, 구성요소들(1520-1590)을 서로 연결하고, 구성요소들 간의 신호(예: 제어 메시지 또는 데이터)를 전달하는 회로를 포함할 수 있다.
- [0203] 프로세서(1520)는, 중앙처리장치(central processing unit, CPU), 어플리케이션 프로세서(application processor, AP), GPU(graphics processing unit), 카메라의 ISP(image signal processor), 또는 CP(communication processor) 중 하나 또는 그 이상을 포함할 수 있다. 일 실시 예에 따르면, 프로세서(1520)는 SoC(system on chip) 또는 SiP(system in package)로 구현될 수 있다. 프로세서(1520)는, 예를 들면, 운영 체제 또는 응용 프로그램을 구동하여 프로세서(1520)에 연결된 전자 장치(1501)의 적어도 하나의 다른 구성요소(예: 하드웨어 또는 소프트웨어 구성요소)를 제어할 수 있고, 각종 데이터 처리 및 연산을 수행할 수 있다. 프로세서(1520)는 다른 구성요소들(예: 통신 모듈(1590)) 중 적어도 하나로부터 수신된 명령 또는 데이터를 휘발성 메모리(1532)에 로드 하여 처리하고, 결과 데이터를 비 휘발성 메모리(1534)에 저장할 수 있다.
- [0204] 메모리(1530)는, 휘발성 메모리(1532) 또는 비 휘발성 메모리(1534)를 포함할 수 있다. 휘발성 메모리(1532)는, 예를 들면, RAM(random access memory)(예: DRAM, SRAM, 또는 SDRAM)로 구성될 수 있다. 비 휘발성 메모리(1534)는, 예를 들면, PROM(programmable read-only memory), OTPROM(one time PROM), EPROM(erasable PROM), EEPROM(electrically EPROM), mask ROM, flash ROM, 플래시 메모리, HDD(hard disk drive), 또는 SSD(solid state drive)로 구성될 수 있다. 또한, 비 휘발성 메모리(1534)는, 전자 장치(1501)와의 연결 형태에 따라, 그 안에 배치된 내장 메모리(1536), 또는 필요 시에만 연결하여 사용 가능한 스탠드-얼론(stand-alone) 형태의 외장 메모리(1538)로 구성될 수 있다. 외장 메모리(1538)는 플래시 드라이브(flash drive), 예를 들면, CF(compact flash), SD(secure digital), Micro-SD, Mini-SD, xD(extreme digital), MMC(multi-media card), 또는 메모리 스틱을 포함할 수 있다. 외장 메모리(1538)는 유선(예: 케이블 또는 USB(universal serial bus)) 또는 무선(예: Bluetooth)을 통하여 전자 장치(1501)와 기능적으로 또는 물리적으로 연결될 수 있다.
- [0205] 메모리(1530)는, 예를 들면, 전자 장치(1501)의 적어도 하나의 다른 소프트웨어 구성요소, 예를 들어, 프로그램(1540)에 관계된 명령 또는 데이터를 저장할 수 있다. 프로그램(1540)은, 예를 들면, 커널(1541), 라이브러리(1543), 어플리케이션 프레임워크(1545), 또는 어플리케이션 프로그램(interchangeably "어플리케이션")(1547)을 포함할 수 있다.
- [0206] 입력 장치(1550)는, 마이크, 마우스, 또는 키보드를 포함할 수 있다. 일 실시 예에 따르면, 키보드는 물리적인 키보드로 연결되거나, 표시 장치(1560)를 통해 가상 키보드로 표시될 수 있다.
- [0207] 표시 장치(1560)는, 디스플레이, 홀로그램 장치, 또는 프로젝터 및 해당 장치를 제어하기 위한 제어 회로를 포함할 수 있다. 디스플레이는, 예를 들면, 액정 디스플레이(LCD), 발광 다이오드(LED) 디스플레이, 유기 발광 다

이오드(OLED) 디스플레이, 마이크로 전자기계 시스템(MEMS) 디스플레이, 또는 전자 종이(electronic paper) 디스플레이를 포함할 수 있다. 디스플레이는, 일 실시 예에 따르면, 유연하게, 투명하게, 또는 착용할 수 있게 구현될 수 있다. 디스플레이는 사용자의 터치, 제스처, 근접, 또는 호버링(hovering) 입력을 감지할 수 터치 회로(touch circuitry) 또는 터치에 대한 압력의 세기를 측정할 수 있는 압력 센서(interchangeably "force sensor")를 포함할 수 있다. 상기 터치 회로 또는 압력 센서는 디스플레이와 일체형으로 구현되거나, 또는 디스플레이와는 별도의 하나 이상의 센서들로 구현될 수 있다. 홀로그램 장치는 빛의 간섭을 이용하여 입체 영상을 허공에 보여줄 수 있다. 프로젝터는 스크린에 빛을 투사하여 영상을 표시할 수 있다. 스크린은, 예를 들면, 전자 장치(1501)의 내부 또는 외부에 위치할 수 있다.

[0208] 오디오 모듈(1570)은, 예를 들면, 소리와 전기 신호를 쌍방향으로 변환시킬 수 있다. 일 실시 예에 따르면, 오디오 모듈(1570)은, 입력 장치(1550)(예: 마이크)를 통해 소리를 획득하거나, 또는 전자 장치(1501)에 포함된 출력 장치(미 도시)(예: 스피커 또는 리시버), 또는 전자 장치(1501)와 연결된 외부 전자 장치(예: 전자 장치(1502)(예: 무선 스피커 또는 무선 헤드폰) 또는 전자 장치(1506)(예: 유선 스피커 또는 유선 헤드폰))를 통해 소리를 출력할 수 있다.

[0209] 센서 모듈(1576)은, 예를 들면, 전자 장치(1501)의 내부의 작동 상태(예: 전력 또는 온도), 또는 외부의 환경 상태(예: 고도, 습도, 또는 밝기)를 측정 또는 감지하여, 그 측정 또는 감지된 상태 정보에 대응하는 전기 신호 또는 데이터 값을 생성할 수 있다. 센서 모듈(1576)은, 예를 들면, 제스처 센서, 자이로 센서, 기압 센서, 마그네틱 센서, 가속도 센서, 그립 센서, 근접 센서, 컬러(color) 센서(예: RGB(red, green, blue) 센서), IR(infrared) 센서, 생체 센서(예: 홍채 센서, 지문 센서, 또는 HRM(heart rate monitoring) 센서, 후각(electronic nose) 센서, EMG(electromyography) 센서, EEG(Electroencephalogram) 센서, ECG(Electrocardiogram) 센서), 온도 센서, 습도 센서, 조도 센서, 또는 UV(ultra violet) 센서를 포함할 수 있다. 센서 모듈(1576)은 그 안에 속한 적어도 하나 이상의 센서들을 제어하기 위한 제어 회로를 더 포함할 수 있다. 어떤 실시 예에서는, 전자 장치(1501)는 프로세서(1520) 또는 프로세서(1520)와는 별도의 프로세서(예: 센서 허브)를 이용하여, 센서 모듈(1576)을 제어할 수 있다. 별도의 프로세서(예: 센서 허브)를 이용하는 경우에, 전자 장치(1501)는 프로세서(1520)가 슬립(sleep) 상태에 있는 동안, 프로세서(1520)를 깨우지 않고 별도의 프로세서의 작동에 의하여 센서 모듈(1576)의 동작 또는 상태의 적어도 일부를 제어할 수 있다.

[0210] 인터페이스(1577)는, 일 실시 예에 따르면, HDMI(high definition multimedia interface), USB, 광 인터페이스(optical interface), RS-232(recommended standard 232), D-sub(D-subminiature), MHL(mobile high-definition link) 인터페이스, SD카드/MMC(multi-media card) 인터페이스, 또는 오디오 인터페이스를 포함할 수 있다. 연결 단자(1578)는 전자 장치(1501)와 전자 장치(1506)를 물리적으로 연결시킬 수 있다. 일 실시 예에 따르면, 연결 단자(1578)는, 예를 들면, USB 커넥터, SD 카드/MMC 커넥터, 또는 오디오 커넥터(예: 헤드폰 커넥터)를 포함할 수 있다.

[0211] 햅틱 모듈(1579)은 전기적 신호를 기계적인 자극(예: 진동 또는 움직임) 또는 전기적인 자극으로 변환할 수 있다. 예를 들면, 햅틱 모듈(1579)은 사용자에게 촉각 또는 운동 감각과 관련된 자극을 제공할 수 있다. 햅틱 모듈(1579)은 예를 들면, 모터, 압전 소자, 또는 전기 자극 장치를 포함할 수 있다.

[0212] 카메라 모듈(1580)은, 예를 들면, 정지 영상 및 동영상을 촬영할 수 있다. 카메라 모듈(1580)은, 일 실시 예에 따르면, 하나 이상의 렌즈(예: 광각 렌즈 및 망원 렌즈, 또는 전면 렌즈 및 후면 렌즈), 이미지 센서, 이미지 시그널 프로세서, 또는 플래시(예: 발광 다이오드 또는 제논 램프(xenon lamp) 등)를 포함할 수 있다.

[0213] 전력 관리 모듈(1588)은 전자 장치(1501)의 전력을 관리하기 위한 모듈로서, 예를 들면, PMIC(power management integrated circuit)의 적어도 일부로서 구성될 수 있다.

[0214] 배터리(1589)는, 예를 들면, 1차 전지, 2차 전지, 또는 연료 전지를 포함하여 외부 전원에 의해 재충전되어, 상기 전자 장치(1501)의 적어도 하나의 구성 요소에 전력을 공급할 수 있다.

[0215] 통신 모듈(1590)은, 예를 들면, 전자 장치(1501)와 외부 장치(예: 제1 외부 전자 장치(1502), 제2 외부 전자 장치(1504), 또는 서버(1508)) 간의 통신 채널 수립 및 수립된 통신 채널을 통한 유선 또는 무선 통신의 수행을 지원할 수 있다. 일 실시 예에 따르면, 통신 모듈(1590)은 무선 통신 모듈(1592) 또는 유선 통신 모듈(1594)을 포함하고, 그 중 해당하는 통신 모듈을 이용하여 제1 네트워크(1598)(예: Bluetooth 또는 IrDA(infrared data association)와 같은 근거리 통신 네트워크) 또는 제2 네트워크(1599)(예: 셀룰러 네트워크와 같은 원거리 통신 네트워크)를 통하여 외부 장치와 통신할 수 있다.

- [0216] 무선 통신 모듈(1592)은, 예를 들면, 셀룰러 통신, 근거리 무선 통신, 또는 GNSS 통신을 지원할 수 있다. 셀룰러 통신은, 예를 들면, LTE(long-term evolution), LTE-A(LTE Advance), CDMA(code division multiple access), WCDMA(wideband CDMA), UMTS(universal mobile telecommunications system), WiBro(Wireless Broadband), 또는 GSM(Global System for Mobile Communications)을 포함할 수 있다. 근거리 무선 통신은, 예를 들면, Wi-Fi(wireless fidelity), Wi-Fi Direct, Li-Fi(light fidelity), Bluetooth, BLE(Bluetooth low energy), Zigbee, NFC(near field communication), MST(magnetic secure transmission), RF(radio frequency), 또는 BAN(body area network)을 포함할 수 있다. GNSS는, 예를 들면, GPS(Global Positioning System), Glonass(Global Navigation Satellite System), Beidou Navigation Satellite System(이하 "Beidou") 또는 Galileo(the European global satellite-based navigation system)을 포함할 수 있다. 본 문서에서 "GPS"는 "GNSS"와 상호 호환적으로 사용될 수 있다.
- [0217] 일 실시 예에 따르면, 상기 무선 통신 모듈(1592)은, 셀룰러 통신을 지원하는 경우, 예를 들면, 가입자 식별 모듈(1596)을 이용하여 통신 네트워크 내에서 전자 장치(1501)의 구별 및 인증을 수행할 수 있다. 일 실시 예에 따르면, 무선 통신 모듈(1592)은 프로세서(1520)(예: AP)와 별개인 CP를 포함할 수 있다. 이런 경우, CP는, 예를 들면, 프로세서(1520)가 인액티브(예: 슬립) 상태에 있는 동안 프로세서(1520)를 대신하여, 또는 프로세서(1520)가 액티브 상태에 있는 동안 프로세서(1520)와 함께, 전자 장치(1501)의 구성요소들(1510-1596) 중 적어도 하나의 구성 요소와 관련된 기능들의 적어도 일부 기능을 수행할 수 있다. 일 실시 예에 따르면, 무선 통신 모듈(1592)은 셀룰러 통신 모듈, 근거리 무선 통신 모듈, 또는 GNSS 통신 모듈 중 해당하는 통신 방식만을 지원하는 복수의 통신 모듈들로 구성될 수 있다.
- [0218] 유선 통신 모듈(1594)은, 예를 들면, LAN(local area network), 전력선 통신 또는 POTS(plain old telephone service)를 포함할 수 있다.
- [0219] 제1 네트워크(1598)는, 예를 들어, 전자 장치(1501)와 제1 외부 전자 장치(1502)간의 무선으로 직접 연결을 통해 명령 또는 데이터를 송신 또는 수신 할 수 있는 Wi-Fi 다이렉트 또는 Bluetooth를 포함할 수 있다. 제2 네트워크(1599)는, 예를 들어, 전자 장치(1501)와 제2 외부 전자 장치(1504)간의 명령 또는 데이터를 송신 또는 수신할 수 있는 텔레커뮤니케이션 네트워크(예: LAN(local area network)나 WAN(wide area network)와 같은 컴퓨터 네트워크, 인터넷(internet), 또는 텔레폰(telephone) 네트워크)를 포함할 수 있다.
- [0220] 다양한 실시 예들에 따르면, 상기 명령 또는 상기 데이터는 제2 네트워크에 연결된 서버(1508)를 통해서 전자 장치(1501)와 제2 외부 전자 장치(1504)간에 송신 또는 수신될 수 있다. 제1 및 제2 외부 전자 장치(1502, 1504) 각각은 전자 장치(1501)와 동일한 또는 다른 종류의 장치일 수 있다. 다양한 실시 예들에 따르면, 전자 장치(1501)에서 실행되는 동작들의 전부 또는 일부는 다른 하나 또는 복수의 전자 장치(예: 전자 장치(1502, 1504), 또는 서버(1508)에서 실행될 수 있다. 일 실시 예에 따르면, 전자 장치(1501)가 어떤 기능이나 서비스를 자동으로 또는 요청에 의하여 수행해야 할 경우에, 전자 장치(1501)는 기능 또는 서비스를 자체적으로 실행시키는 대신에 또는 추가적으로, 그와 연관된 적어도 일부 기능을 다른 장치(예: 전자 장치(1502, 1504), 또는 서버(1508))에게 요청할 수 있다. 다른 전자 장치(예: 전자 장치(1502, 1504), 또는 서버(1508))는 요청된 기능 또는 추가 기능을 실행하고, 그 결과를 전자 장치(1501)로 전달할 수 있다. 전자 장치(1501)는 수신된 결과를 그대로 또는 추가적으로 처리하여 요청된 기능이나 서비스를 제공할 수 있다. 이를 위하여, 예를 들면, 클라우드 컴퓨팅, 분산 컴퓨팅, 또는 클라이언트-서버 컴퓨팅 기술이 이용될 수 있다.
- [0221] 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.
- [0222] 본 문서의 다양한 실시 예들 및 이에 사용된 용어들은 본 문서에 기재된 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 해당 실시 예의 다양한 변경, 균등물, 및/또는 대체물을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다. 본 문서에서, "A 또는 B", "A 및/또는 B 중 적어도 하나", "A, B 또는 C" 또는 "A, B 및/또는 C 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다. "제1," "제2," "첫째," 또는 "둘째," 등의 표현들은 해당 구성요소들을, 순서 또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다. 어떤(예: 제1) 구성요소가 다른(예: 제2) 구성요소에 "(기능적으로 또는 통

신적으로) 연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제 3 구성요소)를 통하여 연결될 수 있다.

[0223] 본 문서에서, "~하도록 설정된(adapted to or configured to)"은 상황에 따라, 예를 들면, 하드웨어적 또는 소프트웨어적으로 "~에 적합한," "~하는 능력을 가지는," "~하도록 변경된," "~하도록 만들어진," "~를 할 수 있는," 또는 "~하도록 설계된"과 상호 호환적으로(interchangeably) 사용될 수 있다. 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 설정된 (또는 구성된) 프로세서"는 해당 동작들을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치(예: 메모리 1530)에 저장된 하나 이상의 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(예: CPU 또는 AP)를 의미할 수 있다.

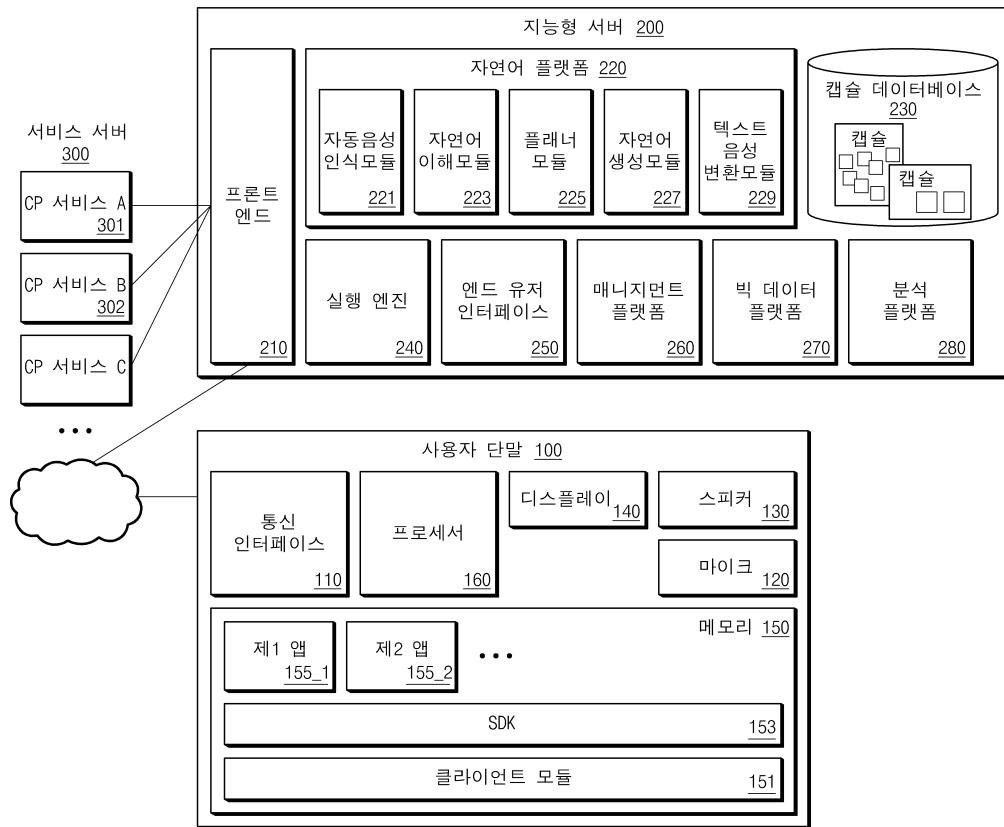
[0224] 본 문서에서 사용된 용어 "모듈"은 하드웨어, 소프트웨어 또는 펌웨어(firmware)로 구성된 유닛(unit)을 포함하며, 예를 들면, 로직, 논리 블록, 부품, 또는 회로 등의 용어와 상호 호환적으로 사용될 수 있다. "모듈"은, 일체로 구성된 부품 또는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수 있다. "모듈"은 기계적으로 또는 전자적으로 구현될 수 있으며, 예를 들면, 어떤 동작들을 수행하는, 알려졌거나 앞으로 개발될, ASIC(application-specific integrated circuit) 칩, FPGAs(field-programmable gate arrays), 또는 프로그램 가능 논리 장치를 포함할 수 있다.

[0225] 다양한 실시 예들에 따른 장치(예: 모듈들 또는 그 기능들) 또는 방법(예: 동작들)의 적어도 일부는 프로그램 모듈의 형태로 컴퓨터로 판독 가능한 저장 매체(예: 메모리(1530))에 저장된 명령어로 구현될 수 있다. 상기 명령어가 프로세서(예: 프로세서(1520))에 의해 실행될 경우, 프로세서가 상기 명령어에 해당하는 기능을 수행할 수 있다. 컴퓨터로 판독 가능한 기록 매체는, 하드디스크, 플로피디스크, 마그네틱 매체(예: 자기테이프), 광기록 매체(예: CD-ROM, DVD, 자기-광 매체(예: 플롭티컬 디스크), 내장 메모리 등을 포함할 수 있다. 명령어는 컴파일러에 의해 만들어지는 코드 또는 인터프리터에 의해 실행될 수 있는 코드를 포함할 수 있다.

[0226] 다양한 실시 예들에 따른 구성 요소(예: 모듈 또는 프로그램 모듈) 각각은 단수 또는 복수의 개체로 구성될 수 있으며, 전술한 해당 서브 구성 요소들 중 일부 서브 구성 요소가 생략되거나, 또는 다른 서브 구성 요소를 더 포함할 수 있다. 대체적으로 또는 추가적으로, 일부 구성 요소들(예: 모듈 또는 프로그램 모듈)은 하나의 개체로 통합되어, 통합되기 이전의 각각의 해당 구성 요소에 의해 수행되는 기능을 동일 또는 유사하게 수행할 수 있다. 다양한 실시 예들에 따른 모듈, 프로그램 모듈 또는 다른 구성 요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱(heuristic)하게 실행되거나, 적어도 일부 동작이 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

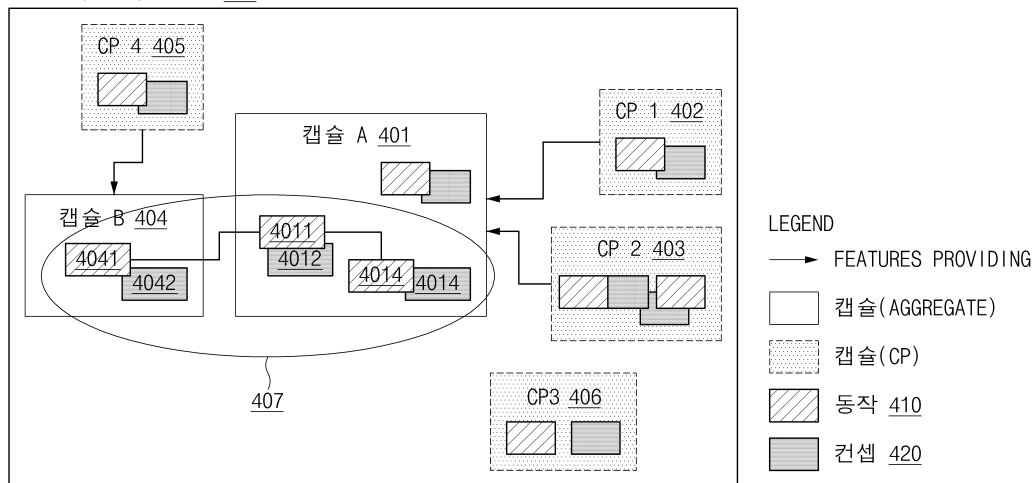
도면

도면1



도면2

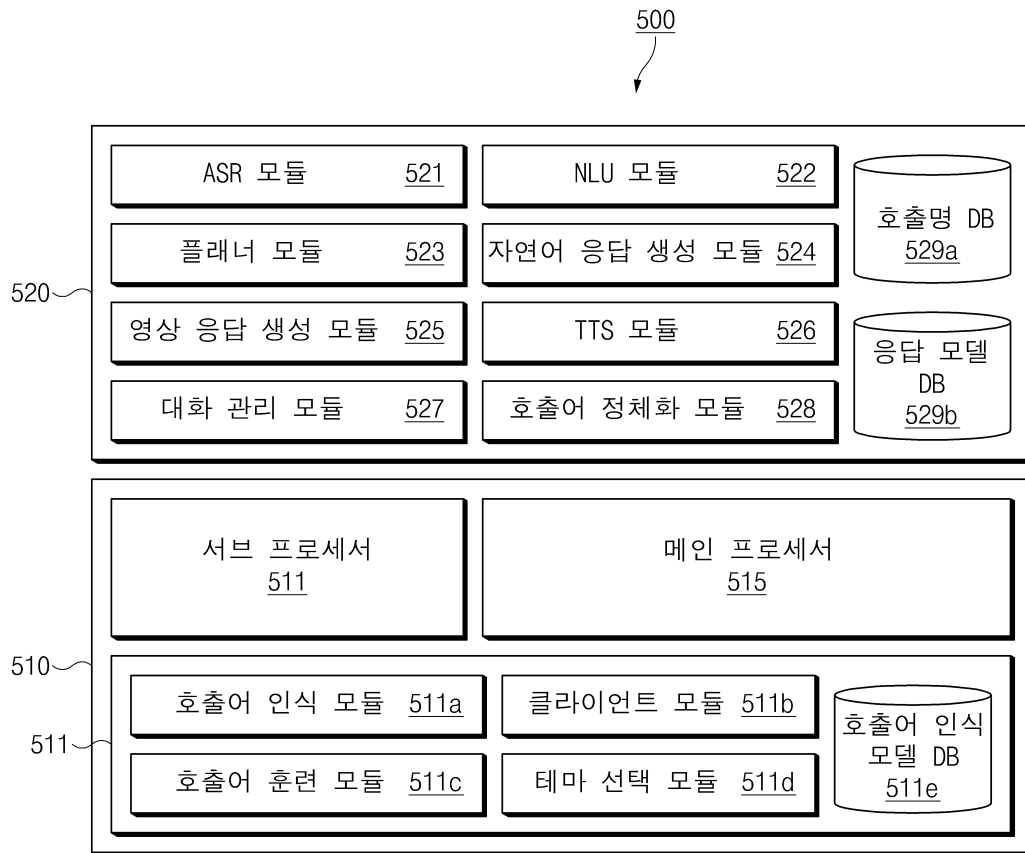
컨셉 액션 네트워크 400



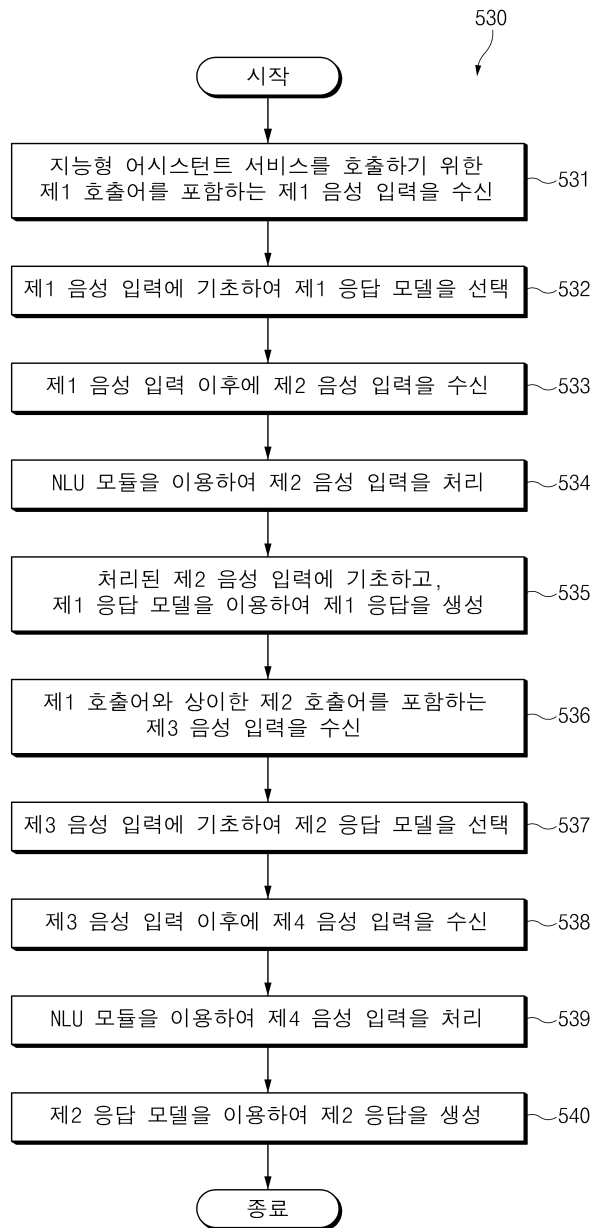
도면3



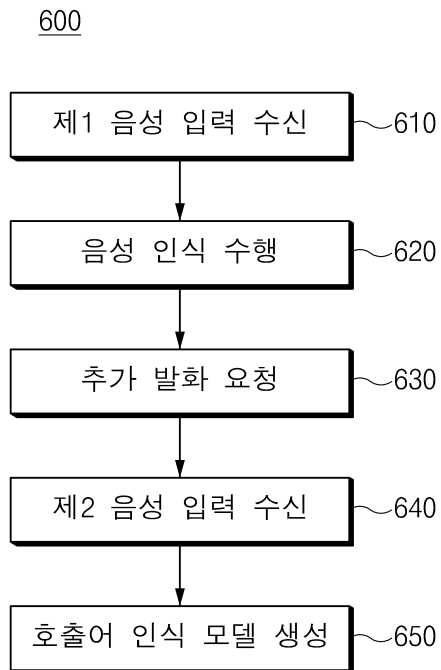
도면4a



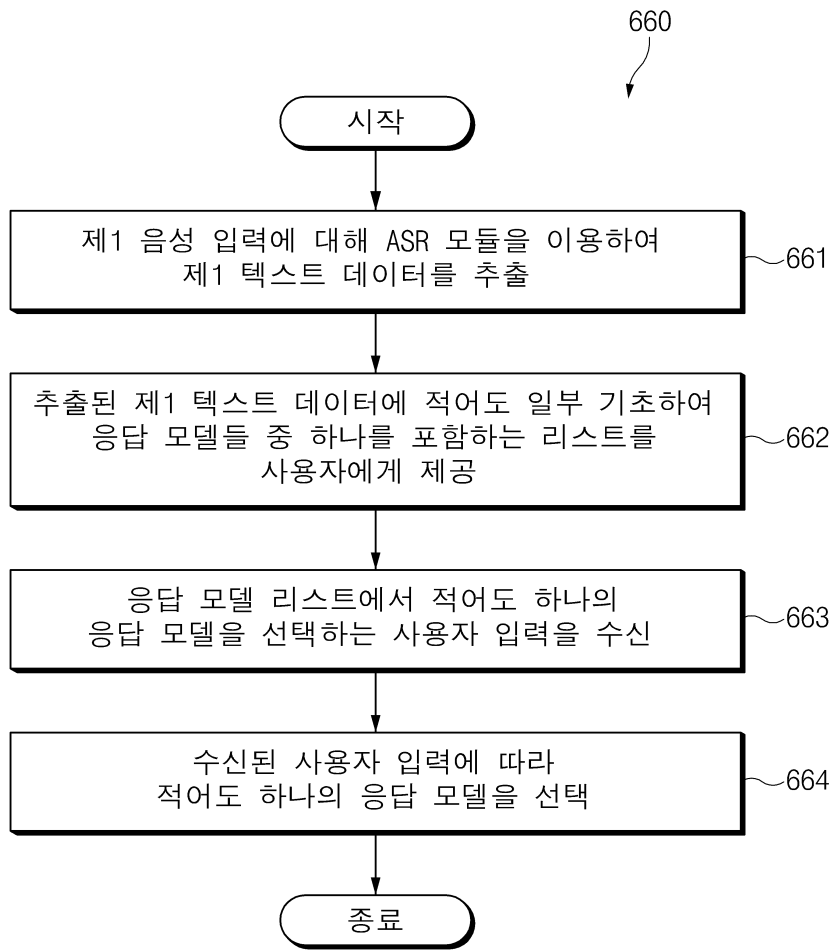
도면4b



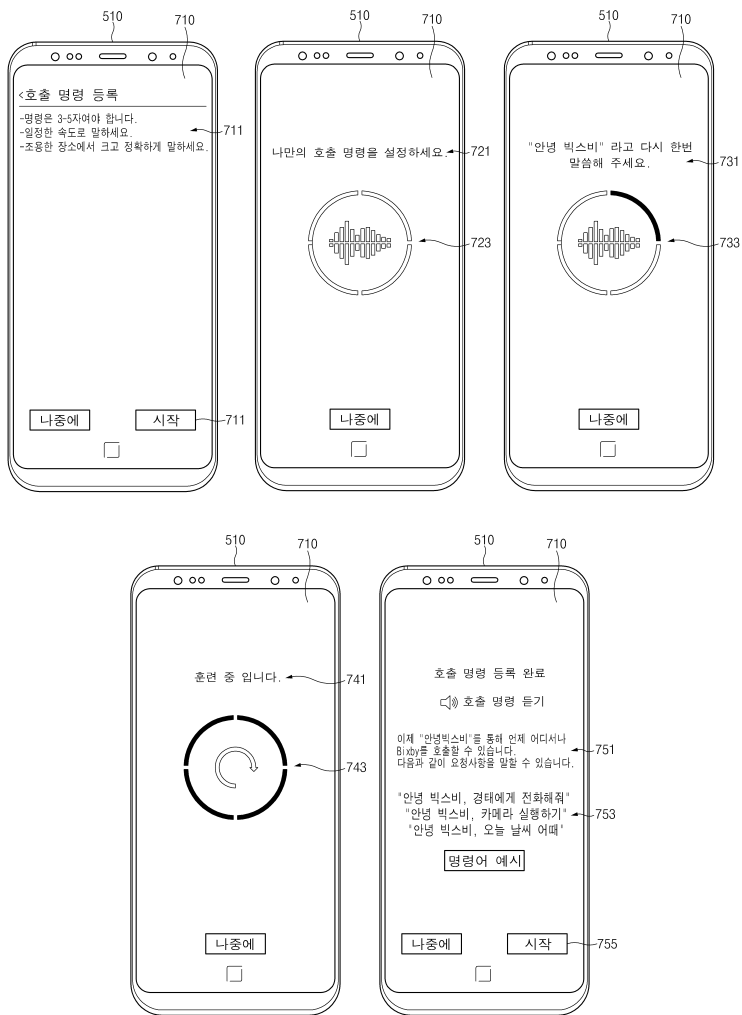
도면5a



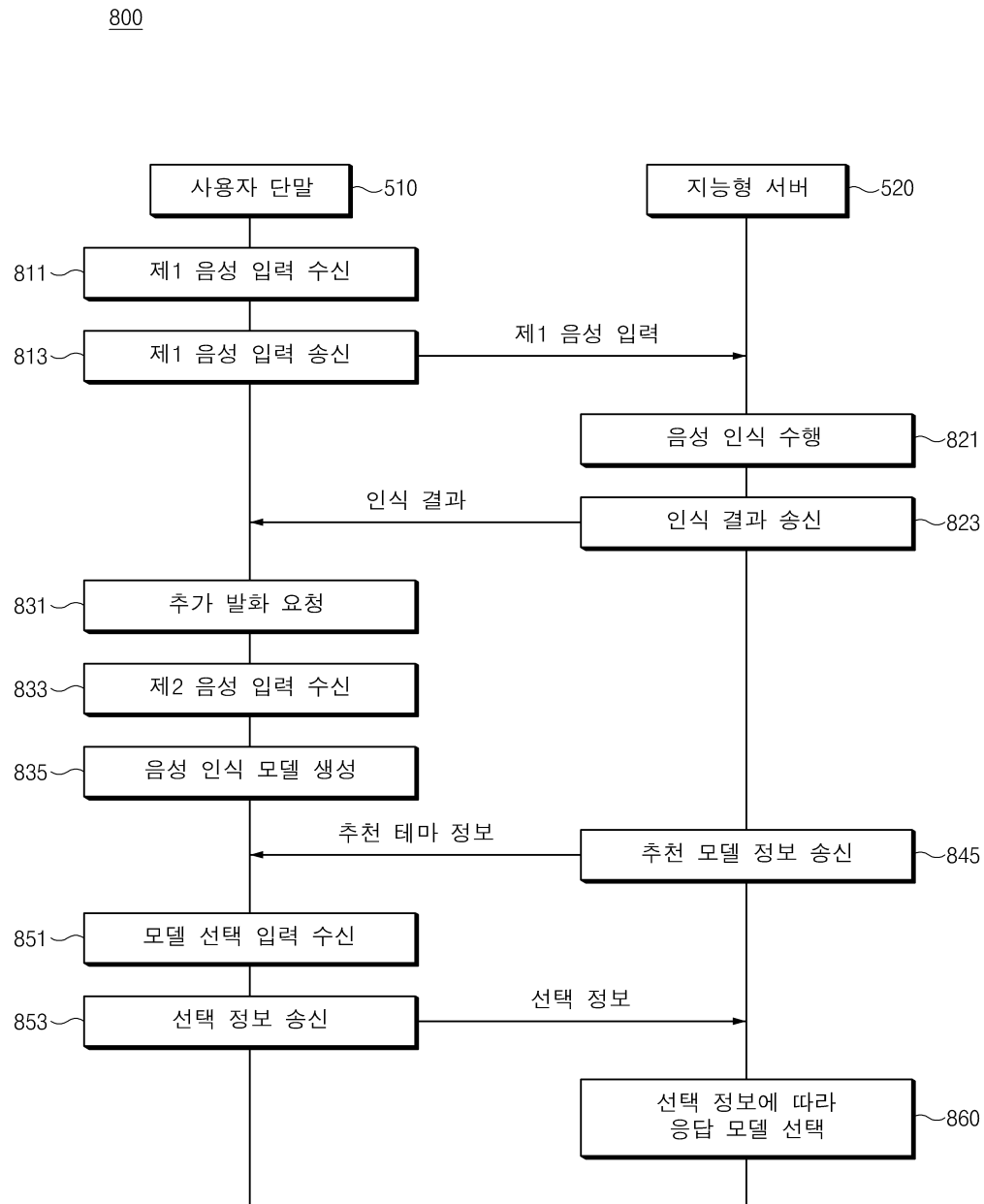
도면5b



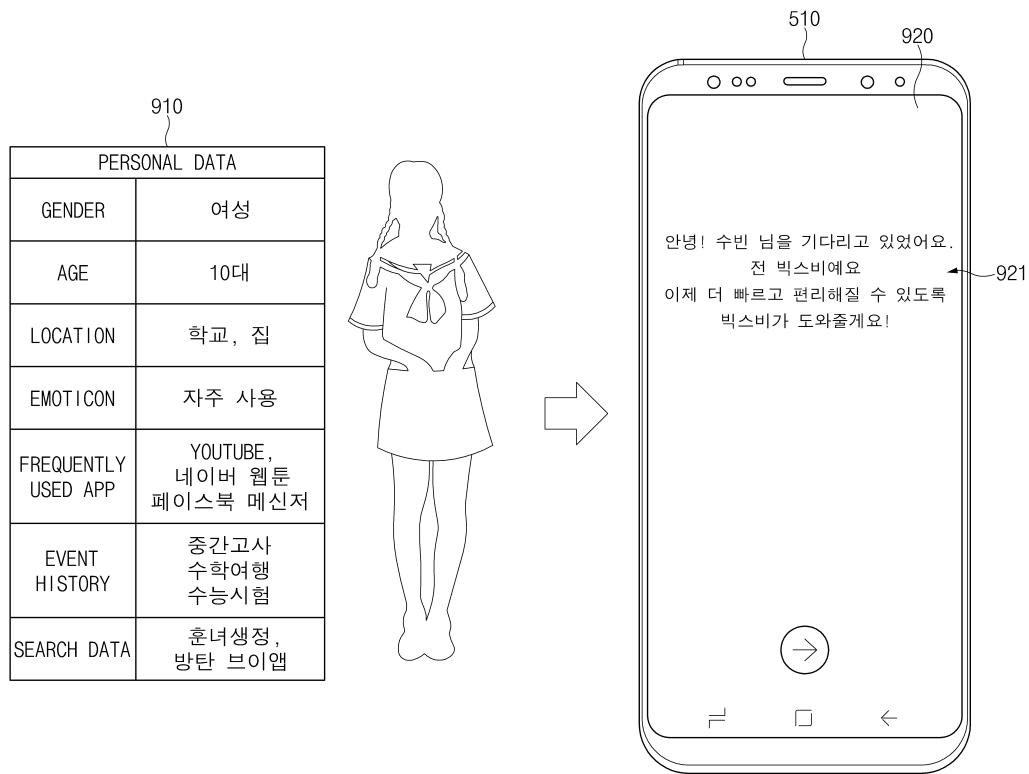
도면6



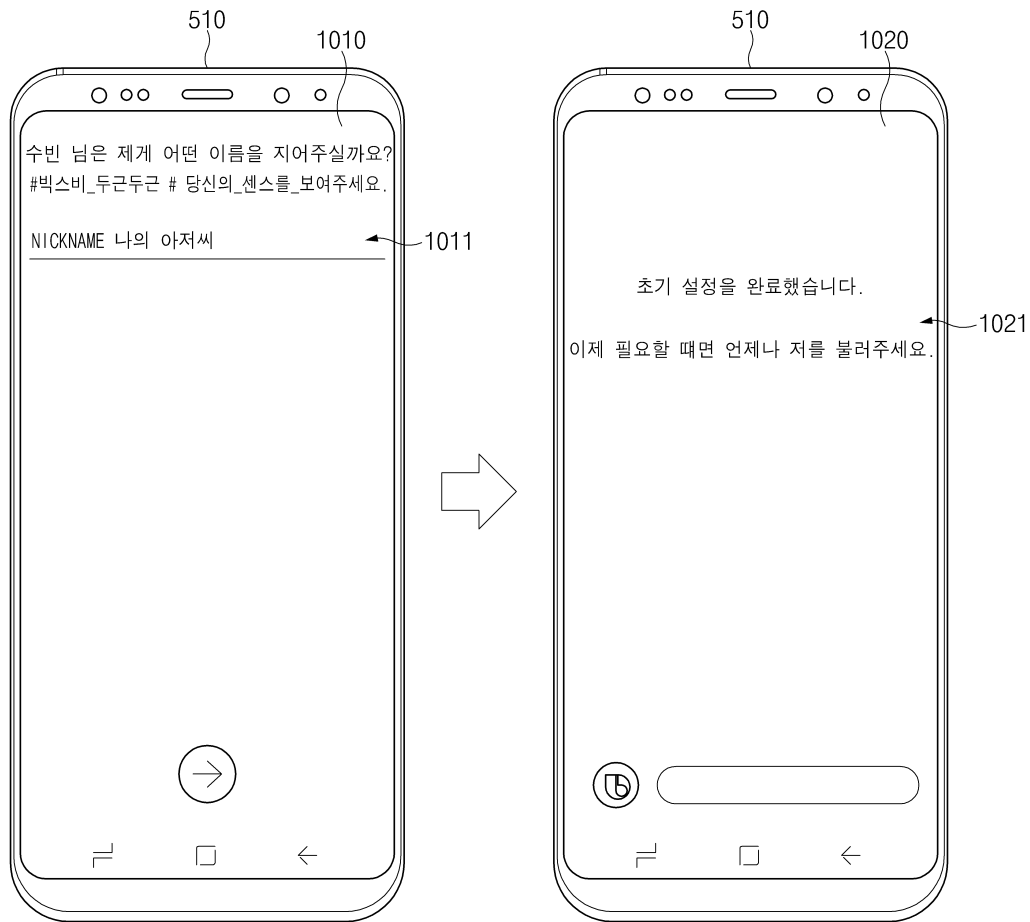
도면7



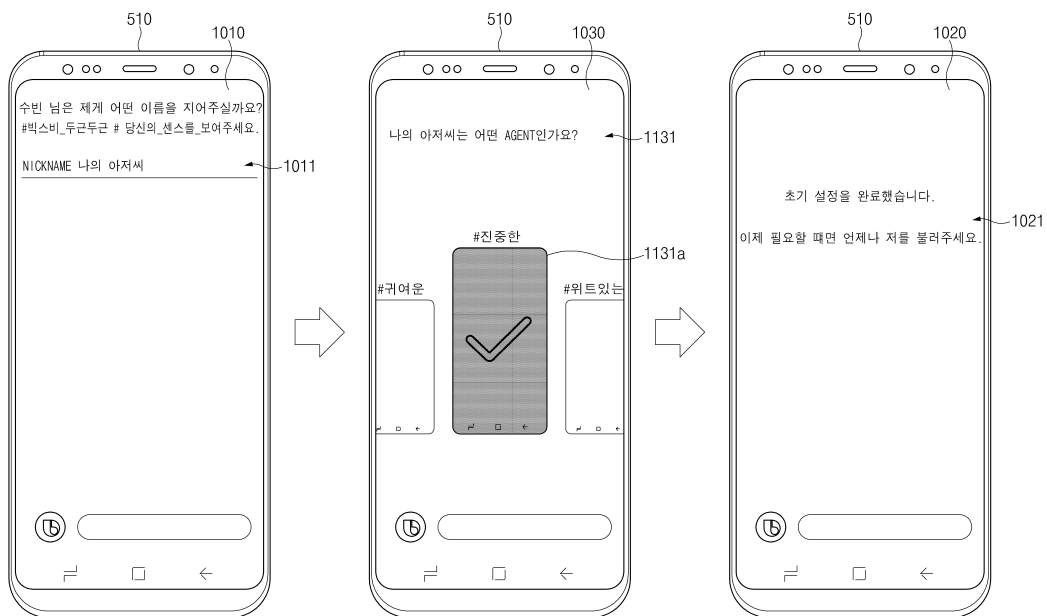
도면8



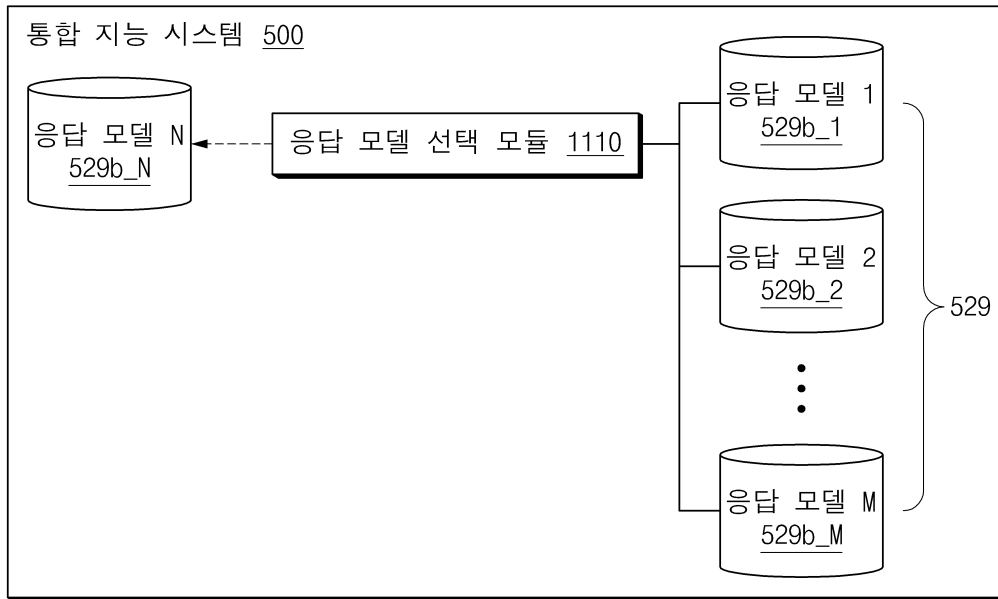
도면9a



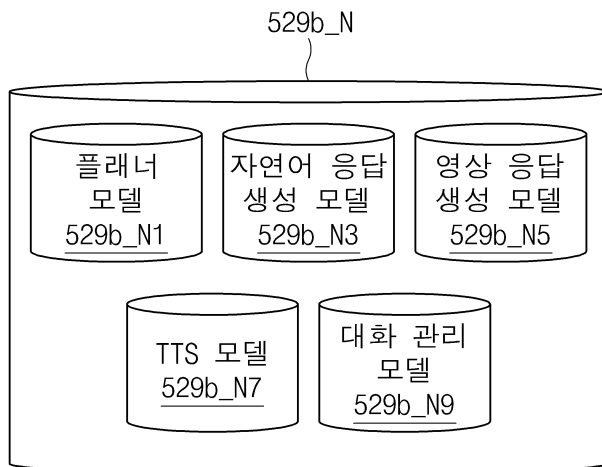
도면9b



도면10a

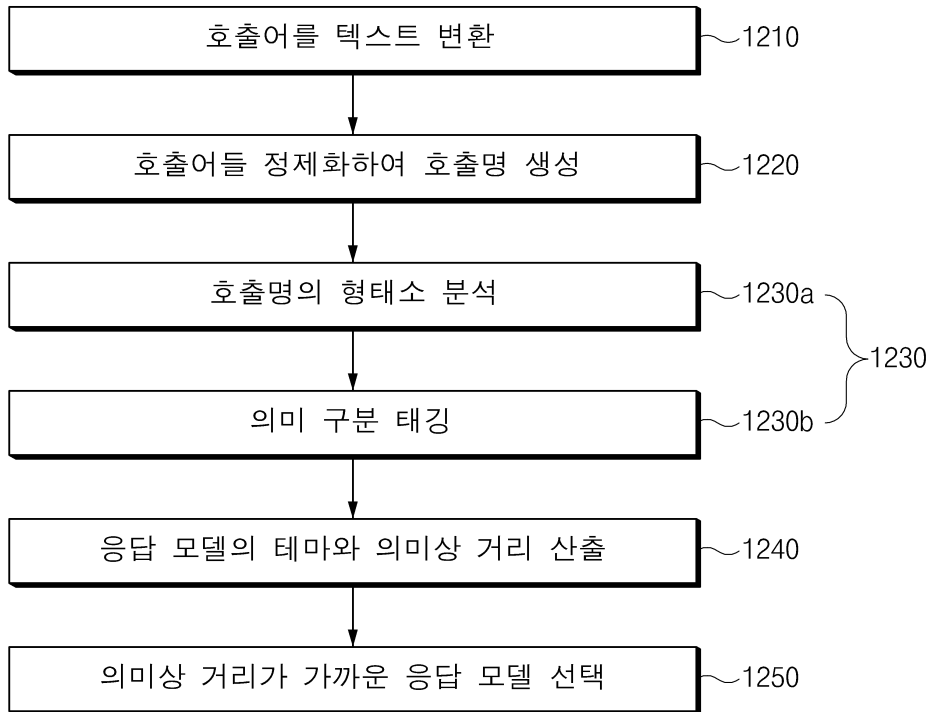


도면10b



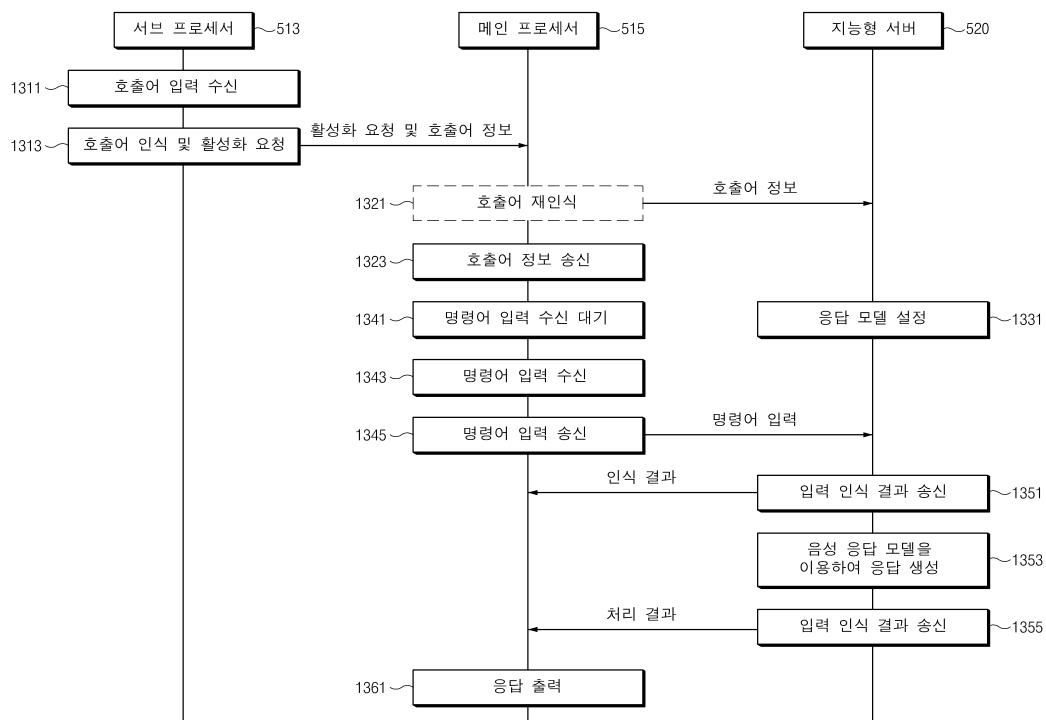
도면11

1200

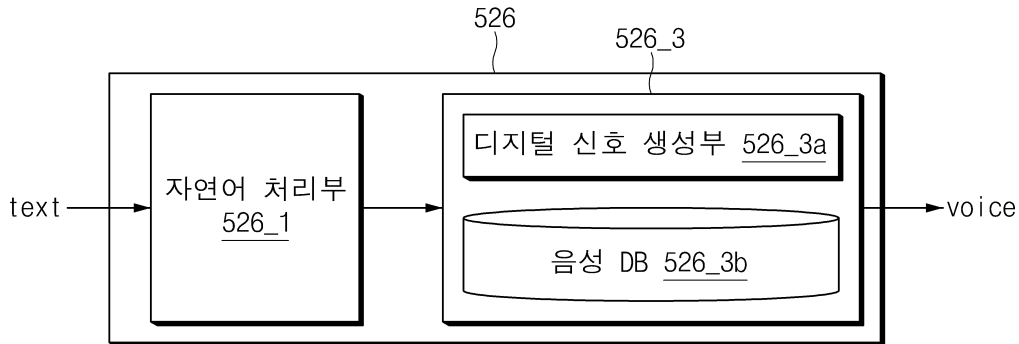


도면12

1300



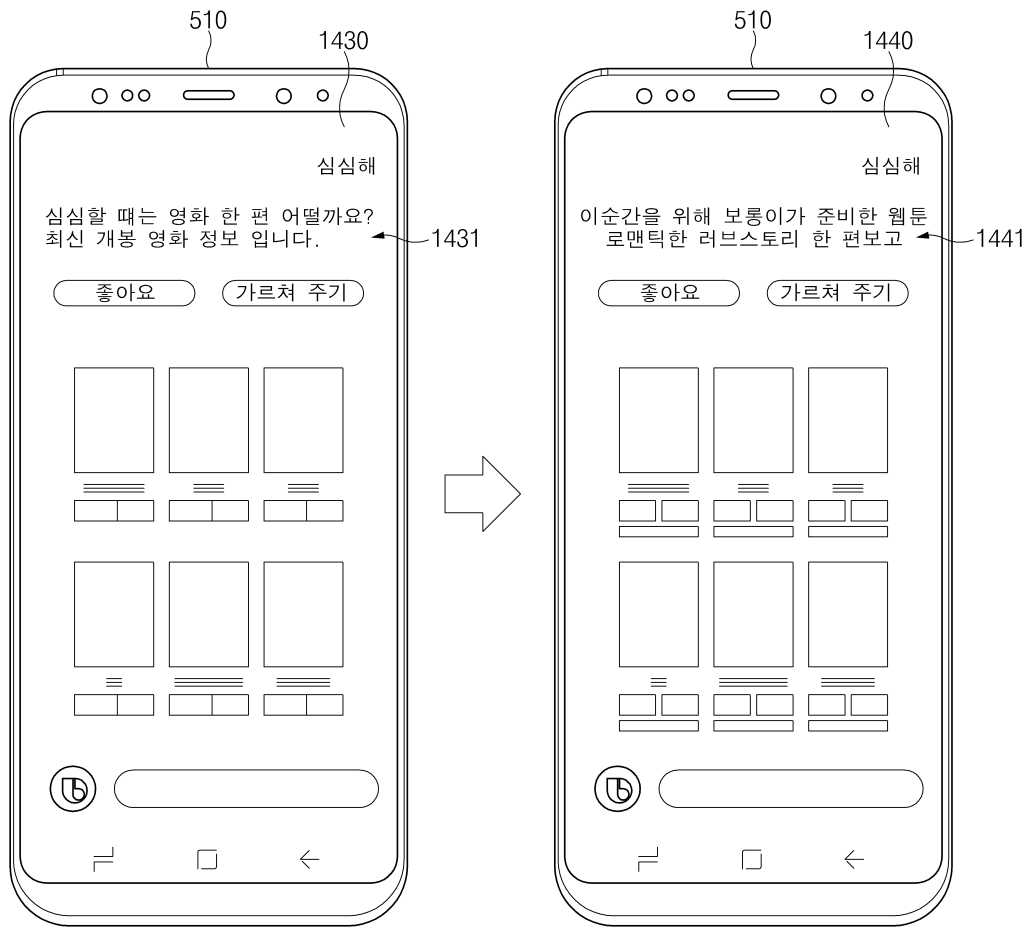
도면13



도면14a



도면14b



도면15

