



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2022년03월07일  
(11) 등록번호 10-2371513  
(24) 등록일자 2022년03월02일

(51) 국제특허분류(Int. Cl.)  
G10L 15/22 (2006.01) G10L 15/04 (2006.01)  
G10L 15/25 (2013.01) G10L 17/02 (2013.01)  
G10L 21/0216 (2013.01)  
(52) CPC특허분류  
G10L 15/22 (2013.01)  
G10L 15/04 (2013.01)  
(21) 출원번호 10-2017-0175608  
(22) 출원일자 2017년12월20일  
심사청구일자 2020년12월16일  
(65) 공개번호 10-2019-0074344  
(43) 공개일자 2019년06월28일  
(56) 선행기술조사문헌  
KR1020150112594 A  
KR1020150027608 A

(73) 특허권자  
현대자동차주식회사  
서울특별시 서초구 현릉로 12 (양재동)  
기아 주식회사  
서울특별시 서초구 현릉로 12 (양재동)  
(72) 발명자  
이정엽  
경기도 용인시 수지구 신봉2로 26 (신봉동, LG신  
봉자이1차아파트) 113동 104호  
신동수  
경기도 수원시 영통구 월드컵로 70 (원천동, 중앙  
기독교초등학교)69 광교호반베르디움(A18) 1004동  
1204호  
김선아  
서울특별시 관악구 승방3나길 13(남현동, 이가채)  
103동 401호  
(74) 대리인  
특허법인세림

전체 청구항 수 : 총 14 항

심사관 : 강민정

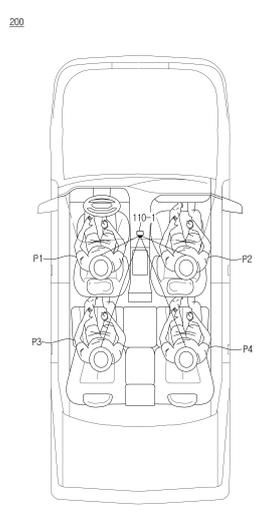
(54) 발명의 명칭 대화 시스템 및 대화 처리 방법

(57) 요약

본 발명의 일 측면은 차량의 다수의 사용자가 존재하는 상황에서 대화 시스템이 어느 시점에 개입하고 어떤 사용자의 말에 응답을 실시하는지 결정하여 효율적인 서비스를 제공하는 대화 시스템 및 대화 처리 방법을 제공한다.

일 실시예에 따른 대화 시스템은, 적어도 한 명의 사용자 정보를 저장하는 저장부; 상기 적어도 한 명의 사용자의 발화 정보를 획득하는 제1센서부; 상기 적어도 한 명의 사용자의 영상 정보를 획득하는 제2센서부; 및 상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고, 상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정하는 대화관리기;를 포함한다.

대표도 - 도3



(52) CPC특허분류

*G10L 15/25* (2013.01)

*G10L 17/02* (2013.01)

*G10L 2015/223* (2013.01)

*G10L 2021/02166* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

적어도 한 명의 사용자 정보를 저장하는 저장부;  
상기 적어도 한 명의 사용자의 발화 정보를 획득하는 제1센서부;  
상기 적어도 한 명의 사용자의 영상 정보를 획득하는 제2센서부; 및  
상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고,  
상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정하는 대화관리기;를 포함하는 대화 시스템.

#### 청구항 2

제1항에 있어서,  
상기 저장부는,  
상기 적어도 한 명의 사용자의 기준 위치 정보를 저장하고,  
상기 대화관리기는,  
상기 발화 정보를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하고,  
상기 현재 위치 정보를 상기 기준 위치 정보에 대응하여 상기 기준 사용자를 판단하는 대화 시스템.

#### 청구항 3

제1항에 있어서,  
상기 저장부는,  
상기 적어도 한 명의 사용자의 기준 음성 정보를 저장하고,  
상기 대화관리기는,  
상기 발화 정보를 상기 기준 음성 정보에 대응하여 상기 기준 사용자를 판단하는 대화 시스템.

#### 청구항 4

제1항에 있어서,  
상기 제1센서부는,  
차량에 내부에 위치하고 빔포밍(Beamforming) 신호를 수신하는 적어도 하나의 마이크를 포함하고,  
상기 대화관리기는,  
상기 적어도 하나의 마이크가 출력하는 상기 빔포밍 신호를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하는 대화 시스템.

**청구항 5**

제1항에 있어서,  
상기 제1센서부는,  
상기 입력 개시 신호의 발화 시점 이전의 상기 적어도 한 명의 사용자의 과거 발화 정보를 획득하고,  
상기 대화관리기는,  
상기 과거 발화 정보를 기초로 상기 사용자 정보를 생성하여 저장하는 대화 시스템.

**청구항 6**

제1항에 있어서,  
상기 대화관리기는,  
상기 영상 정보를 기초로 상기 적어도 한 명의 사용자의 위치 정보를 도출하는 대화 시스템.

**청구항 7**

제1항에 있어서,  
상기 제2센서부는,  
상기 적어도 한 명의 사용자의 입술 주변부를 포함하는 안면 영상 정보를 획득하고,  
상기 대화관리기는,  
상기 적어도 한 명의 사용자의 입술 움직임을 기초로 상기 입력 개시 신호의 발화 여부를 판단하여, 상기 기준 사용자를 판단하는 대화 시스템.

**청구항 8**

적어도 한 명의 사용자 정보를 저장하고,  
상기 적어도 한 명의 사용자의 발화 정보를 획득하고,  
상기 적어도 한 명의 사용자의 영상 정보를 획득하고,  
상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고,  
상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정하는 것을 포함하는 대화 처리 방법.

**청구항 9**

제8항에 있어서,  
상기 적어도 한 명의 사용자 정보를 저장하는 것은,  
상기 적어도 한 명의 사용자의 기준 위치 정보를 저장하는 것을 포함하고,  
상기 기준 사용자를 판단하는 것은,  
상기 발화 정보를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하고,  
상기 현재 위치 정보를 상기 기준 위치 정보에 대응하여 상기 기준 사용자를 판단하는 것을 포함하는 대화 처리

방법.

**청구항 10**

제8항에 있어서,  
상기 적어도 한 명의 사용자 정보를 저장하는 것은,  
상기 적어도 한 명의 사용자의 기준 음성 정보를 저장하는 것을 포함하고,  
상기 기준 사용자를 판단하는 것은,  
상기 발화 정보를 상기 기준 음성 정보에 대응하여 상기 기준 사용자를 판단하는 것을 포함하는 대화 처리 방법.

**청구항 11**

제8항에 있어서,  
상기 적어도 한 명의 사용자의 발화 정보를 획득하는 것은,  
차량에 내부에 위치하고 빔포밍(Beamforming) 신호를 수신하는 것을 포함하고,  
상기 기준 사용자를 판단하는 것은,  
상기 빔포밍 신호를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하는 것을 포함하는 대화 처리 방법.

**청구항 12**

제8항에 있어서,  
상기 적어도 한 명의 사용자의 발화 정보를 획득하는 것은,  
상기 입력 개시 신호의 발화 시점 이전의 상기 적어도 한 명의 사용자의 과거 발화 정보를 획득하는 것을 포함하고,  
상기 적어도 한 명의 사용자 정보를 저장하는 것은,  
상기 과거 발화 정보를 기초로 상기 사용자 정보를 생성하여 저장하는 것을 포함하는 대화 처리 방법.

**청구항 13**

제8항에 있어서,  
상기 영상 정보를 기초로 상기 적어도 한 명의 사용자의 위치 정보를 도출하는 것을 더 포함하는 대화 처리 방법.

**청구항 14**

제8항에 있어서,  
상기 적어도 한 명의 사용자의 영상 정보를 획득하는 것은,  
상기 적어도 한 명의 사용자의 입술 주변부를 포함하는 안면 영상 정보를 획득하는 것을 더 포함하고,  
상기 기준 사용자를 판단하는 것은,

상기 적어도 한 명의 사용자의 입술 움직임을 기초로 상기 입력 개시 신호의 발화 여부를 판단하여, 상기 기준 사용자를 판단하는 것을 포함하는 대화 처리 방법.

**발명의 설명**

**기술 분야**

[0001] 개시된 발명은 사용자와의 대화를 통해 사용자의 의도를 파악하고 사용자에게 필요한 정보나 서비스를 제공하는 대화 시스템, 이를 포함하는 차량 및 대화 처리 방법에 관한 것이다.

**배경 기술**

[0002] 차량용 AVN이나 대부분의 모바일 기기는 작은 화면 및 작은 버튼으로 인해, 사용자에게 시각적인 정보를 제공하거나 사용자의 입력을 수신함에 있어 불편함이 발생할 수 있다.

[0003] 특히, 사용자가 운전 중 시각적인 정보를 확인하거나 기기를 조작하기 위해, 시선을 이동시키고 스티어링 휠에서 손을 떼는 것은 안전 운전에 위협이 되는 요소로 작용한다.

[0004] 한편, 차량 내 여러 명이 탑승하여 하나의 대화 시스템에 명령을 입력 할 경우, 즉 다수의 사용자가 존재하는 상황에서 대화의 주도권이 시간에 따라 변경될 수 있다. 이와 같은 경우 대화 시스템이 어느 시점에 개입하고 누구의 말에 응대를 해야 하는지 판단, 결정하는 기능이 필요하게 된다.

[0005] 한편 이러한 기능을 기초로 여러 명의 탑승자가 존재하는 상황에서도 사용자와의 대화를 통해 사용자의 의도를 파악하고, 사용자에게 필요한 서비스를 제공하는 대화 시스템이 차량에 적용될 경우 보다 안전하고 편리하게 서비스를 제공할 수 있을 것으로 기대된다.

**발명의 내용**

**해결하려는 과제**

[0006] 본 발명의 일 측면은 차량의 다수의 사용자가 존재하는 상황에서 대화 시스템이 어느 시점에 개입하고 어떤 사용자의 말에 응답을 실시하는지 결정하여 효율적인 서비스를 제공하는 대화 시스템 및 대화 처리 방법을 제공한다.

**과제의 해결 수단**

[0007] 일 실시예에 따른 대화 시스템은, 적어도 한 명의 사용자 정보를 저장하는 저장부; 상기 적어도 한 명의 사용자의 발화 정보를 획득하는 제1센서부; 상기 적어도 한 명의 사용자의 영상 정보를 획득하는 제2센서부; 및 상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고, 상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정하는 대화관리기;를 포함한다.

[0008] 상기 저장부는, 상기 적어도 한 명의 사용자의 기준 위치 정보를 저장하고,

[0009] 상기 대화관리기는, 상기 발화 정보를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하고, 상기 현재 위치 정보를 상기 기준 위치 정보에 대응하여 상기 기준 사용자를 판단할 수 있다.

[0010] 상기 저장부는, 상기 적어도 한 명의 사용자의 기준 음성 정보를 저장하고,

[0011] 상기 대화관리기는, 상기 발화 정보를 상기 기준 음성 정보에 대응하여 상기 기준 사용자를 판단할 수 있다.

[0012] 상기 제1센서부는, 차량에 내부에 위치하고 빔포밍(Beamforming) 신호를 수신하는 적어도 하나의 마이크를 포함하고,

[0013] 상기 대화관리기는, 상기 적어도 하나의 마이크가 출력하는 상기 빔포밍 신호를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출할 수 있다.

[0014] 상기 제1센서부는, 상기 입력 개시 신호의 발화 시점 이전의 상기 적어도 한 명의 사용자의 과거 발화 정보를

획득하고,

- [0015] 상기 대화관리기는, 상기 과거 발화 정보를 기초로 상기 사용자 정보를 생성하여 저장할 수 있다.
- [0016] 상기 대화관리기는, 상기 영상 정보를 기초로 상기 적어도 한 명의 사용자의 위치 정보를 도출할 수 있다.
- [0017] 상기 제2센서부는, 상기 적어도 한 명의 사용자의 입술 주변부를 포함하는 안면 영상 정보를 획득하고,
- [0018] 상기 대화관리기는, 상기 적어도 한 명의 사용자의 입술 움직임을 기초로 상기 입력 개시 신호의 발화 여부를 판단하여, 상기 기준 사용자를 판단할 수 있다.
- [0019] 일 실시예에 따른 대화 처리 방법은 적어도 한 명의 사용자 정보를 저장하고,
- [0020] 상기 적어도 한 명의 사용자의 발화 정보를 획득하고, 상기 적어도 한 명의 사용자의 영상 정보를 획득하고, 상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고, 상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정하는 것을 포함한다.
- [0021] 상기 적어도 한 명의 사용자 정보를 저장하는 것은, 상기 적어도 한 명의 사용자의 기준 위치 정보를 저장하는 것을 포함하고,
- [0022] 상기 기준 사용자를 판단하는 것은, 상기 발화 정보를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하고, 상기 현재 위치 정보를 상기 기준 위치 정보에 대응하여 상기 기준 사용자를 판단하는 것을 포함할 수 있다.
- [0023] 상기 적어도 한 명의 사용자 정보를 저장하는 것은, 상기 적어도 한 명의 사용자의 기준 음성 정보를 저장하는 것을 포함하고,
- [0024] 상기 기준 사용자를 판단하는 것은, 상기 발화 정보를 상기 기준 음성 정보에 대응하여 상기 기준 사용자를 판단하는 것을 포함할 수 있다.
- [0025] 상기 적어도 한 명의 사용자의 발화 정보를 획득하는 것은, 차량에 내부에 위치하고 빔포밍(Beamforming) 신호를 수신하는 것을 포함하고,
- [0026] 상기 기준 사용자를 판단하는 것은, 상기 빔포밍 신호를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하는 것을 포함할 수 있다.
- [0027] 상기 적어도 한 명의 사용자의 발화 정보를 획득하는 것은, 상기 입력 개시 신호의 발화 시점 이전의 상기 적어도 한 명의 사용자의 과거 발화 정보를 획득하는 것을 포함하고,
- [0028] 상기 적어도 한 명의 사용자 정보를 저장하는 것은, 상기 과거 발화 정보를 기초로 상기 사용자 정보를 생성하여 저장하는 것을 포함할 수 있다.
- [0029] 일 실시예에 따른 대화 처리 방법은 상기 영상 정보를 기초로 상기 적어도 한 명의 사용자의 위치 정보를 도출하는 것을 더 포함할 수 있다.
- [0030] 상기 적어도 한 명의 사용자의 영상 정보를 획득하는 것은,
- [0031] 상기 적어도 한 명의 사용자의 입술 주변부를 포함하는 안면 영상 정보를 획득하는 것을 더 포함하고,
- [0032] 상기 기준 사용자를 판단하는 것은, 상기 적어도 한 명의 사용자의 입술 움직임을 기초로 상기 입력 개시 신호의 발화 여부를 판단하여, 상기 기준 사용자를 판단하는 것을 포함할 수 있다.

**발명의 효과**

- [0033] 일 실시예에 따른 대화 시스템 및 대화 처리 방법은, 차량의 다수의 사용자가 존재하는 상황에서 대화 시스템이 어느 시점에 개입하고 어떤 사용자의 말에 응답을 실시하는지 결정하여 효율적인 서비스를 제공할 수 있다.

**도면의 간단한 설명**

- [0034] 도1은 일 실시예에 따른 대화 시스템의 제어 블록도이다.
- 도2는 차량 내부의 구성을 나타낸 도면이다.

도3은 일 실시예에 따른 발화 정보를 획득하는 동작을 나타낸 도면이다.

도4는 일 실시예에 따른 빔포밍 신호를 출력하는 동작을 나타낸 도면이다.

도5는 일 실시예에 따른 영상 정보를 획득하는 동작을 나타낸 도면이다.

도6은 일 실시예에 따른 입술의 움직임을 기초로 입력 개시 신호를 도출하는 동작을 설명하기 위한 도면이다.

도7은 일 실시예에 따른 순서도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0035] 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다. 본 명세서가 실시예들의 모든 요소들을 설명하는 것은 아니며, 본 발명이 속하는 기술분야에서 일반적인 내용 또는 실시예들 간에 중복되는 내용은 생략한다. 명세서에서 사용되는 '부, 모듈, 부재, 블록'이라는 용어는 소프트웨어 또는 하드웨어로 구현될 수 있으며, 실시예들에 따라 복수의 '부, 모듈, 부재, 블록'이 하나의 구성요소로 구현되거나, 하나의 '부, 모듈, 부재, 블록'이 복수의 구성요소들을 포함하는 것도 가능하다.
- [0036] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 직접적으로 연결되어 있는 경우뿐 아니라, 간접적으로 연결되어 있는 경우를 포함하고, 간접적인 연결은 무선 통신망을 통해 연결되는 것을 포함한다.
- [0037] 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0038] 단수의 표현은 문맥상 명백하게 예외가 있지 않는 한, 복수의 표현을 포함한다.
- [0039] 각 단계들에 붙여지는 부호는 각 단계들을 식별하기 위해 사용되는 것으로 이들 부호는 각 단계들 상호 간의 순서를 나타내는 것이 아니며, 각 단계들은 문맥상 명백하게 특정 순서를 기재하지 않는 이상 명기된 순서와 다르게 실시될 수 있다.
- [0040] 이하 첨부된 도면들을 참고하여 대화 시스템(100), 이를 포함하는 차량 및 대화 처리 방법의 실시예를 상세하게 설명한다.
- [0041] 일 실시예에 따른 대화 시스템(100)은 사용자의 음성 및 음성 외 입력을 이용하여 사용자의 의도를 파악하고 사용자의 의도에 적합한 서비스 또는 사용자에게 필요한 서비스를 제공하는 장치로서, 서비스 제공의 일 수단 또는 사용자의 의도를 명확히 파악하기 위한 일 수단으로 시스템 발화를 출력함으로써 사용자와 대화를 수행할 수 있다.
- [0042] 당해 실시예에서 사용자에게 제공되는 서비스는 정보의 제공, 차량의 제어, 오디오/비디오/내비게이션 기능의 실행, 외부 서버로부터 가져온 콘텐츠의 제공 등 사용자의 필요나 사용자의 의도에 부응하기 위해 수행되는 모든 동작을 포함할 수 있다.
- [0043] 또한, 일 실시예에 따른 대화 시스템(100)은 차량 환경에 특화된 대화 처리 기술을 제공함으로써, 차량이라는 특수한 환경에서 사용자의 의도를 정확히 파악할 수 있다.
- [0044] 이러한 대화 시스템(100)과 사용자를 연결하는 게이트 웨이는 차량 또는 차량에 연결된 모바일 기기가 될 수 있다. 후술하는 바와 같이, 대화 시스템(100)은 차량에 마련될 수도 있고, 차량 외부의 원격 서버에 마련되어 차량 또는 차량에 연결된 모바일 기기와의 통신을 통해 데이터를 주고 받을 수도 있다.
- [0045] 또한, 대화 시스템(100)의 구성 요소 중 일부는 차량에 마련되고 일부는 원격 서버에 마련되어 대화 시스템(100)의 동작을 차량과 원격 서버에서 부분적으로 수행하는 것도 가능하다.
- [0046] 도1은 일 실시예에 따른 대화 시스템(100)의 제어 블록도이다.
- [0047] 도1을 참조하면, 일 실시예에 따른 대화 시스템(100)은 사용자의 음성 및 음성 외 입력을 포함하는 사용자 입력이나 차량과 관련된 정보 또는 사용자와 관련된 정보를 포함하는 입력을 처리하는 입력 처리기(110), 입력 처리기(110)의 처리 결과를 이용하여 사용자의 의도를 파악하고, 사용자의 의도나 차량의 상태에 대응되는 액션을 결정하는 대화 관리기(120), 대화 관리기(120)의 출력 결과에 따라 특정 서비스를 제공하거나 대화를 계속 이어 나가기 위한 시스템 발화를 출력하는 결과 처리기(130) 및 대화 시스템(100)이 후술하는 동작을 수행하기 위해

필요한 각종 정보를 저장하는 저장부(140)를 포함한다.

- [0048] 입력 처리기(110)는 제1센서부(110-1)와 제2센서부(110-2)를 포함할 수 있다.
- [0049] 제1센서부(110-1)는 발화를 포함한 음성 신호를 입력 마이크와 같은 구성으로 마련될 수 있다.
- [0050] 제1센서부(110-1)는 적어도 한 명의 사용자의 발화 정보를 획득할 수 있다.
- [0051] 또한 제1센서부(110-1)는 차량에 내부에 위치하고 빔포밍(Beamforming) 신호를 출력할 수 있다. 또한 제1센서부(110-1)는 상기 입력 개시 신호의 발화 시점 이전의 상기 적어도 한 명의 사용자의 과거 발화 정보를 획득할 수 있다.
- [0052] 제2센서부(110-2)는 영상 신호를 획득하는 카메라로 마련될 수 있다.
- [0053] 제2센서부(110-2)는 적어도 한 명의 사용자의 영상 정보를 획득할 수 있다.
- [0054] 제2센서부(110-2)는 적어도 한 명의 사용자의 입술 주변부를 포함하는 안면 영상 정보를 획득할 수 있다.
- [0055] 상술한 바와 같이 사용자 음성과 음성 외 입력, 두 가지 종류의 입력을 수신할 수 있다. 음성 외 입력은 사용자의 제스처 인식이나, 입력 장치의 조작을 통해 입력되는 사용자의 음성 외 입력, 차량의 상태를 나타내는 차량 상태 정보, 차량의 주행 환경과 관련된 주행 환경 정보, 사용자의 상태를 나타내는 사용자 정보 등을 포함할 수 있다. 이러한 정보들 외에도 차량과 사용자와 관련된 정보로서, 사용자의 의도를 파악하거나 사용자에게 서비스를 제공하기 위해 사용될 수 있는 정보이면, 모두 입력 처리기(110)의 입력이 될 수 있다. 사용자는 운전자와 탑승자를 모두 포함할 수 있다.
- [0056] 입력 처리기(110)는 입력된 사용자 음성을 인식하여 텍스트 형태의 발화문으로 변환하고, 사용자의 발화문에 자연어 이해(Natural Language Understanding) 기술을 적용하여 사용자의 의도를 파악한다.
- [0057] 또한, 입력 처리기(110)는 사용자 음성 외에 차량의 상태나 주행 환경과 관련된 정보를 수집하고, 수집된 정보를 이용하여 상황을 이해한다.
- [0058] 입력 처리기(110)는 자연어 이해를 통해 파악한 사용자의 의도와 상황에 관련된 정보 등을 대화 관리기(120)로 전달한다. 입력처리기(110)는 사용자의 발화를 입력 받을 수 있다. 발화는 후술하는 바와 같이 음성입력장치(210)를 통해 입력 받을 수 있다. 또한 입력 처리기(110)는 사용자의 발화시 사용자의 피드백 반응을 입력 받을 수 있다. 피드백 반응은 발화와 같은 음성 반응일 수 있으며 음성 외의 반응일 수 있다. 사용자의 피드백 반응이 음성 반응인 경우 발화와 같이 음성입력장치(210)를 통하여 피드백 반응을 입력 받을 수 있다. 또한 사용자의 피드백 반응이 음성 외 반응인 경우 음성 외 입력장치(220)를 통하여 사용자의 피드백 반응을 입력 받을 수 있다. 입력 처리기(110)는 이러한 피드백 반응을 대화관리기(120)로 전달 할 수 있다.
- [0059] 대화 관리기는 상기 사용자 정보, 상기 발화 정보 및 상기 영상 정보 중 적어도 하나를 기초로, 상기 적어도 한 명의 사용자 중 입력 개시 신호를 발화하는 기준 사용자를 판단하고, 상기 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정할 수 있다.
- [0060] 대화 관리기는 상기 발화 정보를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출하고, 상기 현재 위치 정보를 상기 기준 위치 정보에 대응하여 상기 기준 사용자를 판단할 수 있다. 대화관리기는 발화 정보를 상기 기준 음성 정보에 대응하여 상기 기준 사용자를 판단할 수 있다.
- [0061] 대화관리기는 상기 적어도 하나의 마이크가 출력하는 상기 빔포밍 신호를 기초로 상기 적어도 한 명의 사용자의 현재 위치 정보를 도출할 수 있다.
- [0062] 대화관리기는, 상기 과거 발화 정보를 기초로 상기 사용자 정보를 생성하여 저장할 수 있다.
- [0063] 대화관리기는 영상 정보를 기초로 상기 적어도 한 명의 사용자의 위치 정보를 도출할 수 있다.
- [0064] 대화관리기는 적어도 한 명의 사용자의 입술 움직임을 기초로 상기 입력 개시 신호의 발화 여부를 판단하여, 상기 기준 사용자를 판단할 수 있다.
- [0065] 대화 관리기(120)는 입력 처리기(110)로부터 전달된 사용자의 의도, 상황에 관련된 정보 등에 기초하여 사용자의 의도와 현재 상황에 대응되는 액션을 결정하고, 해당 액션을 수행하기 위해 필요한 인자들을 관리한다.
- [0066] 당해 실시예에서 액션은 특정 서비스를 제공하기 위해 수행되는 모든 동작을 의미할 수 있으며, 액션의 종류는 미리 정의될 수 있다. 경우에 따라, 서비스의 제공과 액션의 수행은 동일한 의미일 수 있다. 특히, 대화 관리기

(120)는 입력 처리기(110)로부터 입력 받은 피드백 반응을 기초로 학습하고, 사용자의 의도를 파악 할 수 있다.

- [0067] 예를 들어, 도메인/액션 추론 규칙 DB에 길 안내, 차량 상태 점검, 주유소 추천 등과 같은 액션이 미리 정의될 수 있고, 저장된 추론 규칙에 따라 사용자의 발화에 대응되는 액션, 즉 사용자가 의도하는 액션을 미리 정의된 액션 중에서 추출할 수 있다.
- [0068] 액션의 종류에 대해서는 제한을 두지 않으며, 대화 시스템(100)이 차량(200) 또는 모바일 기기(400)를 통해 수행 가능한 것으로서, 미리 정의되고, 그 추론 규칙이나 다른 액션/이벤트와의 관계 등이 저장되어 있으면 액션이 될 수 있다.
- [0069] 대화 관리자(120)는 결정된 액션에 관한 정보를 결과 처리기(130)로 전달한다.
- [0070] 결과 처리기(130)는 전달된 액션을 수행하기 위해 필요한 대화 응답 및 명령어를 생성하여 출력한다. 대화 응답은 텍스트, 이미지 또는 오디오로 출력될 수 있고, 명령어가 출력되면 출력된 명령어에 대응되는 차량 제어, 외부 콘텐츠 제공 등의 서비스가 수행될 수 있다.
- [0071] 저장부(140)는 적어도 한 명의 사용자 정보를 저장할 수 있다. 저장부(140)는 적어도 한 명의 사용자의 기준 위치 정보를 저장할 수 있다. 저장부(140)는 상기 적어도 한 명의 사용자의 기준 음성 정보를 저장할 수 있다.
- [0072] 저장부(140)는 대화 처리 및 서비스 제공에 필요한 각종 정보를 저장한다. 예를 들어, 자연어 이해에 사용되는 도메인, 액션, 화행, 개체명과 관련된 정보를 미리 저장할 수 있고, 입력된 정보로부터 상황을 이해하기 위해 사용되는 상황 이해 테이블을 저장할 수도 있으며, 차량에 마련된 센서가 감지한 데이터, 사용자와 관련된 정보, 액션 수행을 위해 필요한 정보를 미리 저장할 수도 있다. 특히 저장부(140)는 사용자의 발화에 대응되는 사용자의 피드백 정보를 저장 할 수 있다. 저장부(140)에 저장되는 정보들에 관한 더 자세한 설명은 후술하도록 한다.
- [0073] 저장부(140)는 캐쉬, ROM(Read Only Memory), PROM(Programmable ROM), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM) 및 플래쉬 메모리(Flash memory)와 같은 비휘발성 메모리 소자 또는 RAM(Random Access Memory)과 같은 휘발성 메모리 소자 또는 하드디스크 드라이브(HDD, Hard Disk Drive), CD-ROM과 같은 저장 매체 중 적어도 하나로 구현될 수 있으나 이에 한정되지는 않는다. 저장부(140)는 제어부와 관련하여 전술한 프로세서와 별개의 칩으로 구현된 메모리일 수 있고, 프로세서와 단일 칩으로 구현될 수도 있다.
- [0074] 전술한 바와 같이, 대화 시스템(100)은 차량 환경에 특화된 대화 처리 기술을 제공한다. 대화 시스템(100)의 구성요소가 전부 차량에 포함될 수도 있고, 일부만 포함될 수도 있다. 대화 시스템(100)은 원격 서버에 마련되고 차량은 대화 시스템(100)과 사용자 사이의 게이트웨이의 역할만 할 수도 있다. 어느 경우이던지, 대화 시스템(100)은 차량 또는 차량과 연결된 모바일 기기를 통해 사용자와 연결될 수 있다.
- [0075] 도1에 도시된 대화 시스템(100)의 구성 요소들의 성능에 대응하여 적어도 하나의 구성요소가 추가되거나 삭제될 수 있다. 또한, 구성 요소들의 상호 위치는 시스템의 성능 또는 구조에 대응하여 변경될 수 있다는 것은 당해 기술 분야에서 통상의 지식을 가진 자에게 용이하게 이해될 것이다.
- [0076] 한편, 도1에서 도시된 각각의 구성요소는 소프트웨어 및/또는 Field rogrammable Gate Array(FPGA) 및 주문형 반도체(ASIC, Application Specific Integrated Circuit)와 같은 하드웨어 구성요소를 의미한다.
- [0077] 도2는 차량 내부의 구성을 나타낸 도면이다.
- [0078] 도2를 참조하면, 차량(200) 내부의 대시보드(201)의 중앙 영역인 센터페시아(203)에는 오디오 기능, 비디오 기능, 내비게이션 기능 또는 전화 걸기 기능을 포함하는 차량의 제어를 수행하기 위해 필요한 화면을 표시하는 디스플레이(231)와 사용자의 제어 명령을 입력 받기 위한 입력 버튼(221)이 마련될 수 있다.
- [0079] 또한, 운전자의 조작 편의성을 위해 스티어링 휠(207)에도 입력 버튼(223)이 마련될 수 있고, 운전석(254a)과 조수석(254b) 사이의 센터 콘솔 영역(202)에 입력 버튼의 역할을 수행하는 조그 셔틀(225)이 마련될 수도 있다.
- [0080] 디스플레이(231), 입력 버튼(221) 및 각종 기능을 전반적으로 제어하는 프로세서를 포함하는 모듈을 AVN(Audio Video Navigation) 단말이라 할 수도 있고, 헤드유닛(Head Unit)이라 할 수도 있다.
- [0081] 디스플레이(231)는 LCD(Liquid Crystal Display), LED(Light Emitting Diode), PDP(Plasma Display Panel), OLED(Organic Light Emitting Diode), CRT(Cathode Ray Tube) 등의 다양한 디스플레이 장치 중 하나로 구현될

수 있다.

- [0082] 입력 버튼(221)은 도 2에 도시된 바와 같이 디스플레이(231)와 인접한 영역에 하드 키 타입으로 마련될 수도 있고, 디스플레이(231)가 터치 스크린 타입으로 구현되는 경우에는 디스플레이(231)가 입력 버튼(221)의 기능도 함께 수행할 수 있다.
- [0083] 차량(1)은 사용자의 명령을 음성 입력 장치(210)를 통해 음성으로 입력 받을 수 있다. 음성 입력 장치(210)는 음향을 입력 받아 전기적인 신호로 변환하여 출력하는 마이크를 포함할 수 있다.
- [0084] 사용자 입력 중 음성을 제외한 입력은 음성 외 입력 장치(220)를 통해 입력될 수 있다. 음성 외 입력 장치(220)는 사용자의 조작을 통해 명령을 입력 받는 입력 버튼(221, 223)과 조그 셔틀(225)을 포함할 수 있다.
- [0085] 효과적인 음성의 입력을 위하여 음성 입력 장치(210)는 도 2에 도시된 바와 같이 헤드라이닝(205)에 마련될 수 있으나, 차량(200)의 실시예가 이에 한정되는 것은 아니며, 대시보드(201) 위에 마련되거나 스티어링 휠(207)에 마련되는 것도 가능하다. 이 외에도 사용자의 음성을 입력 받기에 적합한 위치이면 어디든 제한이 없다.
- [0086] 차량(200) 내부에는 사용자와 대화를 수행하거나, 사용자가 원하는 서비스를 제공하기 위해 필요한 음향을 출력하는 스피커(232)가 마련될 수 있다. 일 예로, 스피커(232)는 운전석 도어(253a) 및 조수석 도어(253b) 내측에 마련될 수 있다.
- [0087] 스피커(232)는 내비게이션 경로 안내를 위한 음성, 오디오/비디오 콘텐츠에 포함되는 음향 또는 음성, 사용자가 원하는 정보나 서비스를 제공하기 위한 음성, 사용자의 발화에 대한 응답으로서 생성된 시스템 발화 등을 출력할 수 있다.
- [0088] 일 실시예에 따른 대화 시스템(100)은 차량 환경에 특화된 대화 처리 기술을 이용하여 사용자의 라이프스타일에 최적화된 서비스를 제공하고, 커넥티드카(Connected Car), 사물인터넷(IoT), 인공지능(AI) 등의 기술을 이용한 새로운 서비스를 구성할 수 있다. 또한 대화 시스템(100)은 사용자의 피드백 반응에 기초한 피드백 정보를 기초로 학습할 수 있고 학습한 결과를 기초로 사용자의 의도를 파악 할 수 있다.
- [0089] 일 실시예에 따른 대화 시스템(100)과 같이 차량 환경에 특화된 대화 처리 기술을 적용할 경우, 운전자의 직접 주행 상황에서, 주요 상황(Context)에 대한 인지 및 대응이 용이하다. 유량 부족, 졸음 운전 등 주행에 영향을 미치는 요소에 가중치를 부여하여 서비스를 제공할 수 있고, 대부분의 경우 목적지로 이동 중인 조건을 기반으로 주행 시간, 목적지 정보 등 서비스 제공을 위해 필요한 정보를 용이하게 획득할 수 있다.
- [0090] 또한, 운전자의 의도를 파악하고, 기능을 제안하는 지능적인 서비스를 용이하게 구현할 수 있다. 이는 운전자의 직접 주행 상황에 있어 실시간 정보와 액션을 우선적으로 고려하기 때문이다. 일 예로, 주행 중에 운전자가 주유소를 검색하면, 이는 지금 주유소에 가겠다는 운전자의 의도로 해석될 수 있다. 그러나, 차량이 아닌 환경에서 주유소를 검색할 경우 지금 주유소를 가겠다는 의도 외에 위치 정보 조회, 전화번호 조회, 가격 조회 등 다른 다양한 의도로도 해석되는 것이 가능하다.
- [0091] 또한, 차량은 한정적인 공간이지만, 그 안에 다양한 상황이 존재할 수 있다. 예를 들어, 렌터카 등 생소한 인터페이스의 운전 상황, 대리 운전을 맡긴 상황, 세차 등 차량 관리 상황, 유아를 탑승시키는 상황, 특정 목적지를 찾아가는 상황 등에서 대화 시스템(100)을 활용할 수 있다.
- [0092] 또한, 차량 점검 단계, 출발 준비 단계, 주행 단계, 주차 단계 등 차량의 주행과 그 전후를 구성하는 각각의 단계에서도 다양한 서비스의 기회와 대화 상황들이 발생한다. 특히, 차량 문제의 대처 방안을 모르는 상황, 차량과 각종 외부 기기 연동 상황, 연비 등 주행 습관 확인 상황, 스마트 크루즈 컨트롤(Smart Cruise Control) 등의 안전 지원 기능 활용 상황, 내비게이션 조작 상황, 졸음 운전 상황, 매일 동일한 경로를 반복해서 주행하는 상황, 주정차가 가능한 곳인지 확인해야 하는 상황 등에서 대화 시스템(100)을 활용할 수 있다.
- [0093] 또한 차량(200)은 에어컨시설(222)를 구비 하여 차량 내부의 온도를 적절하게 유지할 수 있다.
- [0094] 카메라는 차량내부에 마련될 수 있으며, 적어도 한 명의 사용자의 영상 정보를 획득할 수 있다. 카메라는 CCD(Charge-Coupled Device) 카메라 또는 CMOS 컬러 이미지 센서를 포함할 수 있다. 여기서 CCD 및 CMOS는 모두 카메라의 렌즈를 통해 들어온 빛을 전기 신호로 바꾸어 저장하는 센서를 의미한다. 구체적으로 CCD(Charge-Coupled Device) 카메라는 전하 결합 소자를 사용하여 영상을 전기 신호로 변환하는 장치이다. 또한, CIS(CMOS Image Sensor)는 CMOS 구조를 가진 저소비, 저전력형의 촬상소자를 의미하며, 디지털 기기의 전자 필름 역할을 수행한다. 일반적으로 CCD는 CIS보다 감도가 좋아 차량(1)에 많이 쓰이지만, 반드시 이에 한정되는 것은

아니다. 즉, 개시된 발명에서 카메라는 위치와 장치에 제한이 없으며, 차량(1)에 탑승한 사용자의 영상 정보를 획득하면 충분하다.

- [0095] 도3은 일 실시예에 따른 발화 정보를 획득하는 동작을 나타낸 도면이다.
- [0096] 도3을 참고하면, 차량에 구비된 대화 시스템(100)은 제1센서부(110-1)가 마련될 수 있다. 전술한 바와 같이 제1센서부(110-1)는 마이크로 구비되어 적어도 한 명의 사용자(P1 내지 P4)의 발화 정보를 획득할 수 있다. 한편, 차량에 마련된 제1센서부(110-1)는 적어도 한 명의 사용자가 입력 개시 신호를 획득할 수 있다.
- [0097] 한편, 사용자 상호간은 대화를 나눌 수 있고, 사용자간 대화 도중 대화 시스템(100)에 신호를 입력하기 위한 입력 개시 신호를 발화 할 수 있다. 예를 들어 한 사용자(P1)는 '잘 들어!', '자! 000야!'와 같은 발화를 통하여 입력 개시 신호를 발화 할 수 있다. 사용자(P1)이 상술한 바와 같은 입력 개시 신호를 발화한다면, 대화 관리기는 상기 사용자를 기준 사용자로 결정할 수 있다. 기준 사용자는 적어도 한 명의 사용자 중 대화 관리기에 직접적인 명령을 입력하는 사용자로, 대화관리기가 기준 사용자를 결정하게 되면, 대화 관리기는 기준 사용자의 발화를 기초로 차량이 수행할 액션을 결정하게 된다.
- [0098] 한편 제1센서부(110-1)는 입력 개시 신호 발화 전에도 사용자들(P1 내지 P4)의 발화 정보를 획득할 수 있고, 대화 관리기는 이를 기초로 사용자 정보를 생성할 수 있다.
- [0099] 한편, 도3에서 설명한 입력 개시 신호는 사용자에게 따라서 미리 결정될 수 있으며 언어의 종류나 형태의 제한은 없다.
- [0100] 도4는 일 실시예에 따른 빔포밍 신호를 출력하는 동작을 나타낸 도면이다.
- [0101] 도4를 참고하면, 마이크로 마련된 제1센서부(110-1)가 빔포밍 신호를 수신하는 것을 나타내고 있다.
- [0102] 빔포밍(Beamforming)은 일정한 배열을 이루고 있는 송/수신기를 이용해서 무선 신호를 보내거나 받아들일 때 방향성을 부여하기 위한 신호처리 기술로, 신호를 받아들이는 경우에는 각 수신기에 도착하는 신호의 위상차에 의해 특정 방향의 신호는 서로 상쇄되고 다른 방향의 신호는 서로 보강되는 성질을 이용하는데, 여러 수신기에 도착하는 신호들을 그대로 받아들여 합하기만 하는 것이 아니라 가중치와 지연시간을 조절한 후에 합하면 원하는 방향의 신호만을 상대적으로 크게 받아들일 수 있는 것을 마이크에 적용한 것이다. 즉 마이크를 포함한 제1센서부(110-1)는 적어도 한 명의 사용자(P1 내지 P4)의 발화 신호를 수신하는데 있어 발화 신호에 방향성을 부여할 수 있다. 즉 제1센서부(110-1)가 수신하는 발화 신호의 방향성을 인식하여 신호를 수신할 수 있다. 제1센서부(110-1)가 방향성을 갖고 발화 신호를 수신하면, 대화 시스템(100)은 각각의 발화 신호의 형성된 위치를 기초로 사용자의 위치를 검출할 수 있다.
- [0103] 예를 들어 사용자(P3)의 발화 신호(S3)에 기초하여 방향성을 갖는 경우 제1센서부(110-1)는 사용자(P3)의 발화 신호를 수신하고 대화관리기는 이와 관련된 방향성을 도출하여 사용자(P3)의 위치 정보를 도출할 수 있다. 대화 시스템(100)은 상술한 방법으로 도출한 사용자의 위치 정보와 저장부(140)에 저장된 사용자 정보를 대응시켜 위치 정보에 대응되는 사용자를 기준 사용자로 결정할 수 있다. 대화 시스템(100)이 결정한 기준 사용자의 발화에 포함된 적어도 하나의 명령어를 기초로 차량이 수행할 액션을 결정할 수 있다.
- [0104] 한편 빔포밍으로 사용자의 위치를 결정하는 동작은 상술한 동작에 한정되지 아니하며, 한 사용자(P3)가 아닌 다른 사용자(P1,P2,P4)도 기준 사용자로 결정될 수 있다.
- [0105] 도5는 일 실시예에 따른 영상 정보를 획득하는 동작을 나타낸 도면이고, 도6은 일 실시예에 따른 입술의 움직임을 기초로 입력 개시 신호를 도출하는 동작을 설명하기 위한 도면이다.
- [0106] 도5를 참고하면, 제2센서부(110-2)는 상술한 바와 같이 영상을 획득할 수 있는 구성으로 마련될 수 있다. 제2센서부(110-2)는 차량에 탑승한 적어도 한 명의 영상 정보를 획득할 수 있다. 영상 정보를 통하여 대화 관리기는 사용자의 제스처 인식, 표정을 포함한 정보를 획득할 수 있다. 예를 들어 제2센서부(110-2)가 사용자(P3)가 손으로 부채질을 하는 제스처를 취하고 발화하는 영상 정보를 획득한 경우 해당 사용자가 차량에 마련된 에어컨을 제어하고자 하는 의도인 것으로 파악할 수 있으므로, 대화 시스템(100)은 이를 기초로 해당 사용자(P3)를 기준 사용자로 결정할 수 있다.
- [0107] 도6은 도5에서 더 나아가 대화 시스템(100)이 획득한 영상을 기초로 독화(Lip reading)를 수행하는 것을 설명하기 위한 도면이다. 사용자가 발화하는데 있어서 입술의 모양을 달리할 수 있다. 이를 기초로 사용자의 발화 여부 및 대화 시스템(100)의 명령여부를 판단할 수 있다.

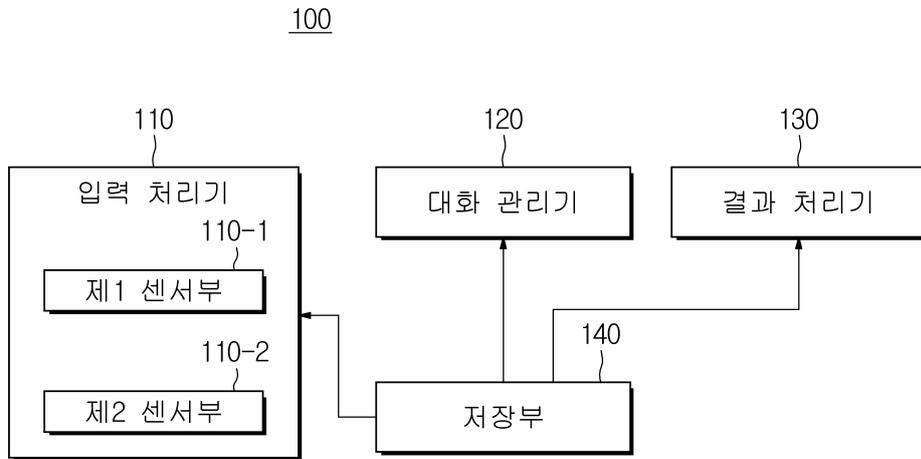
- [0108] 구체적으로 제2센서부(110-2)는 사용자의 안면 영상을 획득할 수 있으며, 안면 영상에는 입술 주변부를 포함할 수 있다. 대화 시스템(100)은 획득한 영상을 기초로 입술 주변부의 특징점(C)을 도출할 수 있다. 대화 시스템(100)은 추출한 특징점의 움직임을 기초로 사용자의 발화 여부를 판단할 수 있다. 즉, 사용자의 발화는 마이크를 포함하고 있는 제1센서부(110-1)가 획득한 발화 정보로 판단할 수 있지만, 제2센서부(110-2)가 획득한 영상 정보로 판단할 수도 있으며, 더 나아가 영상에 포함된 사용자의 입술 주변부의 영상을 기초로 판단 가능하다. 예를 들어 사용자가 '잘들어!'와 같이 발화한 경우 대화 시스템(100)은 영상 정보에 포함된 입술 주변부의 안면 영상의 특징점을 분석하여 사용자의 발화 여부를 판단할 수 있고, 이를 기초로 해당 사용자가 기준사용자인 것으로 결정할 수 있다.
- [0109] 한편 도5및 도6에서 설명한 획득한 영상정보를 해석하여 사용자를 결정하는 동작은 본 발명의 일 실시예에 불과하며 영상 정보를 해석하여 사용자의 발화 여부를 분석하는 동작의 제한은 없다.
- [0110] 도7은 일 실시예에 따른 순서도이다.
- [0111] 도7을 참고하면, 대화 시스템(100)은 발화 정보 및 영상 정보를 획득할 수 있다(1001). 이를 기초로 적어도 한 명의 사용자가 입력 개시 신호를 발화하였는지 여부를 판단할 수 있다(1002). 이를 기초로 입력 개시 신호를 발화한 사용자를 기준 사용자로 판단 할 수 있다(1003). 이후 기준 사용자가 발화한 명령어를 기초로 차량이 수행할 액션을 결정할 수 있다.(1004).
- [0112] 한편, 개시된 실시예들은 컴퓨터에 의해 실행 가능한 명령어를 저장하는 기록매체의 형태로 구현될 수 있다. 명령어는 프로그램 코드의 형태로 저장될 수 있으며, 프로세서에 의해 실행되었을 때, 프로그램 모듈을 생성하여 개시된 실시예들의 동작을 수행할 수 있다. 기록매체는 컴퓨터로 읽을 수 있는 기록매체로 구현될 수 있다.
- [0113] 컴퓨터가 읽을 수 있는 기록매체로는 컴퓨터에 의하여 해독될 수 있는 명령어가 저장된 모든 종류의 기록 매체를 포함한다. 예를 들어, ROM(Read Only Memory), RAM(Random Access Memory), 자기 테이프, 자기 디스크, 플래쉬 메모리, 광 데이터 저장장치 등이 있을 수 있다.
- [0114] 이상에서와 같이 첨부된 도면을 참조하여 개시된 실시예들을 설명하였다.본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 기술적 사상이나 필수적인 특징을 변경하지 않고도, 개시된 실시예들과 다른 형태로 본 발명이 실시될 수 있음을 이해할 것이다. 개시된 실시예들은 예시적인 것이며, 한정적으로 해석되어서는 안 된다.

**부호의 설명**

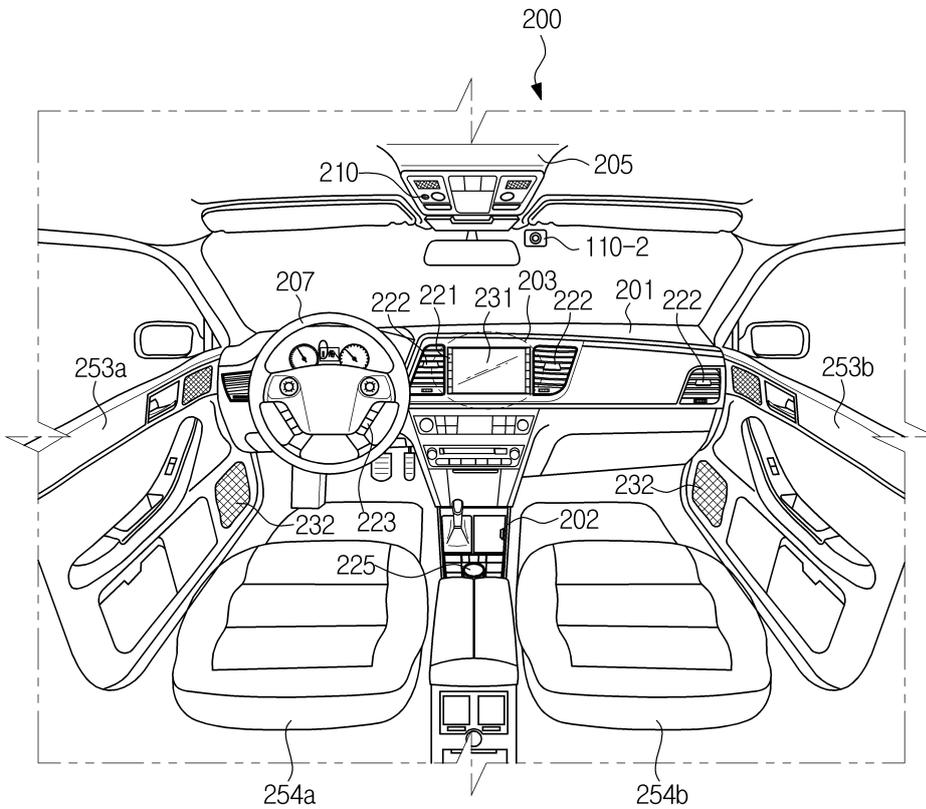
- [0115] 100: 대화 시스템
- 110: 입력 처리기
- 110-1 : 제1센서부
- 110-2 : 제2센서부
- 120: 대화 관리기
- 130: 결과 처리기
- 140 : 저장부
- 200: 차량

도면

도면1

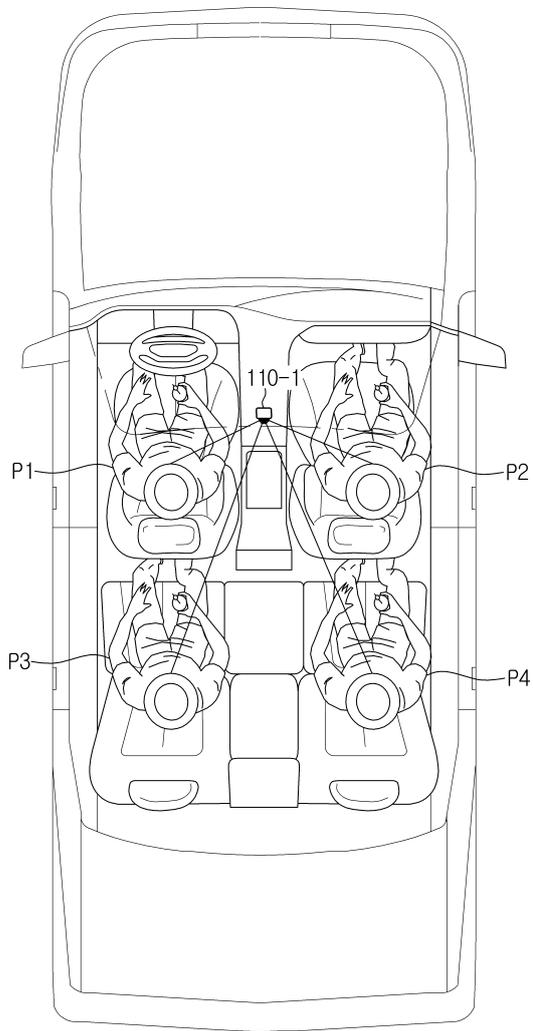


도면2



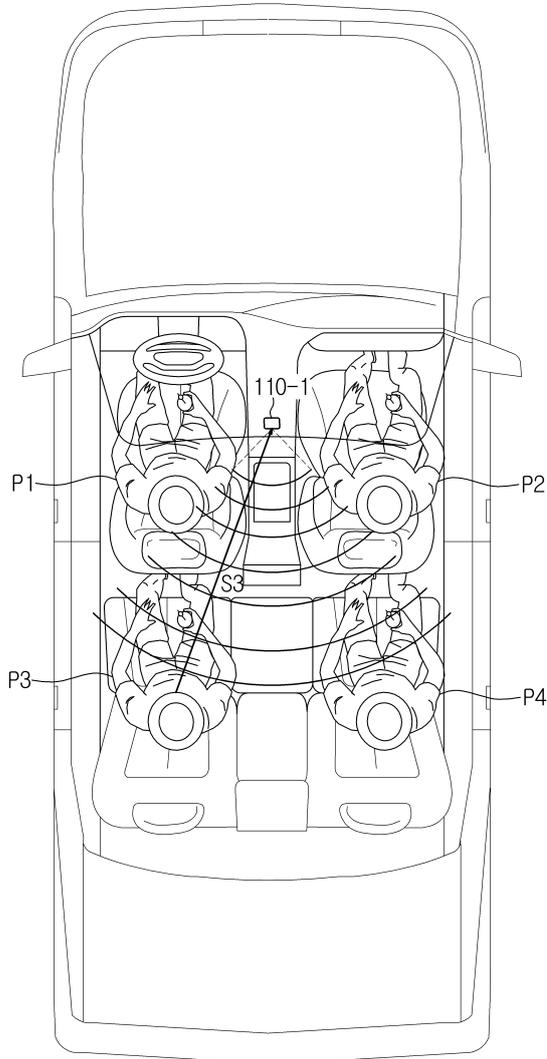
도면3

200

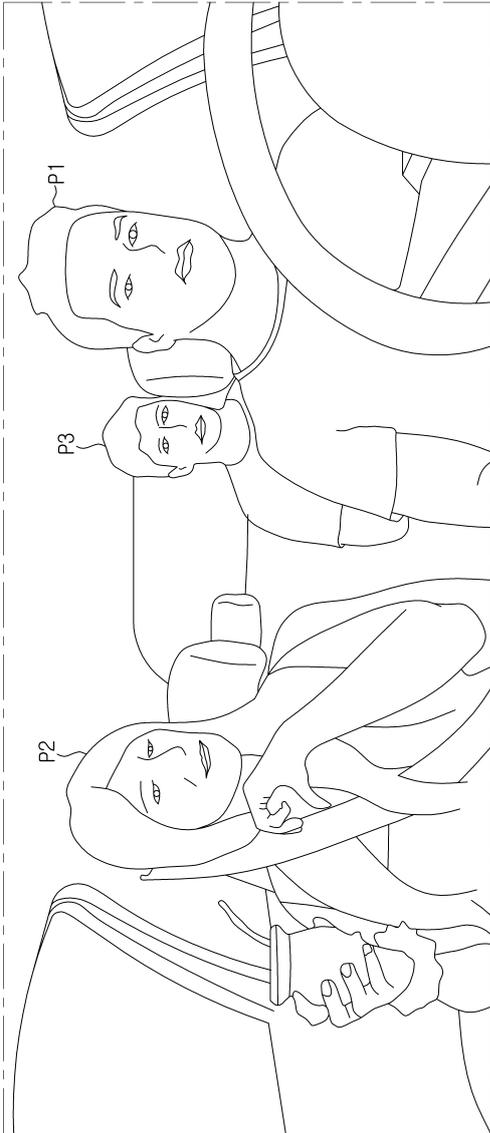


도면4

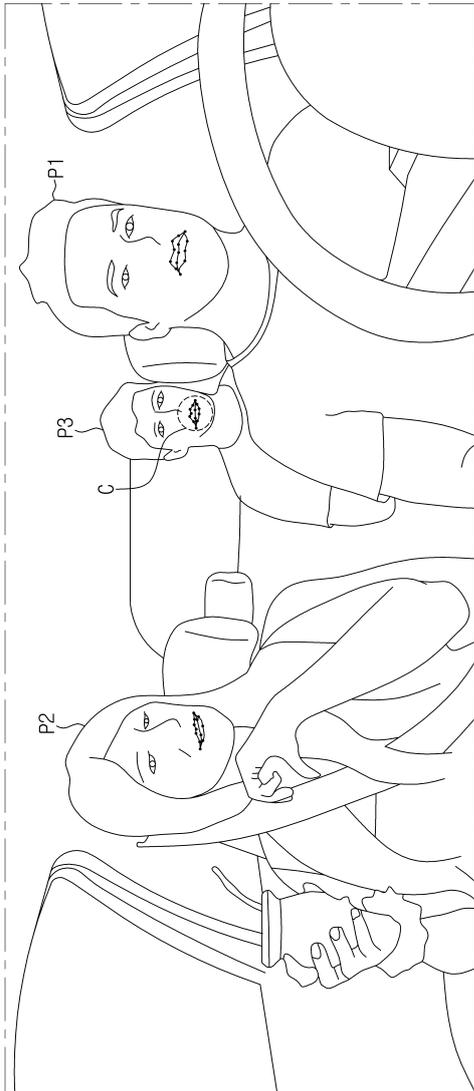
200



도면5



도면6



도면7

