



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2022년04월07일  
(11) 등록번호 10-2384165  
(24) 등록일자 2022년04월04일

(51) 국제특허분류(Int. Cl.)  
HO4N 7/14 (2006.01) G02B 27/01 (2006.01)  
HO4N 7/15 (2016.01)  
(52) CPC특허분류  
HO4N 7/141 (2013.01)  
G02B 27/01 (2013.01)  
(21) 출원번호 10-2020-0129544  
(22) 출원일자 2020년10월07일  
심사청구일자 2020년10월07일  
(65) 공개번호 10-2021-0042028  
(43) 공개일자 2021년04월16일  
(30) 우선권주장  
1020190124882 2019년10월08일 대한민국(KR)  
(56) 선행기술조사문헌  
KR1020180092757 A\*  
KR1020190092326 A\*  
KR1020190095849 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
주식회사 이너테인먼트  
경기도 성남시 분당구 대왕판교로645번길 12 ,5  
층(삼평동, 경기창조경제혁신센터)  
(72) 발명자  
허민강  
서울특별시 성동구 왕십리로31나길 22, 3동 805호  
(74) 대리인  
이영훈, 신무연

전체 청구항 수 : 총 6 항

심사관 : 강석제

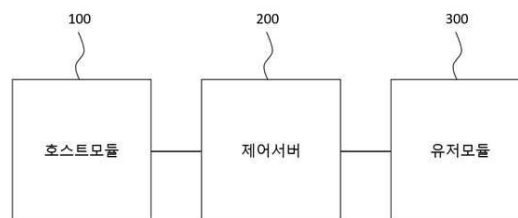
(54) 발명의 명칭 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법 및 시스템

(57) 요약

본 발명은 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법 및 시스템에 있어서, 호스트가 제공하는 영상 또는 콘텐츠가 재생되고 있을 때, 호스트와 유저간에 커뮤니케이션이 가능하고 호스트가 복수의 유저와의 커뮤니케이션을 놓치지 않고 수행할 수 있게 하는 가상공간 내 실시간 양방향 커뮤니케이션 방법 및 시스템에 관한 것으로 호스트에게 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저의 음성을 출력하는 호스트모듈; 상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 적어도 하나 이상의 유저모듈에 전송하고, 상기 유저의 음성 데이터를 수신해 상기 호스트모듈에 전송하는 제어서버; 및 상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 상기 유저에게 출력하고, 상기 유저가 입력한 음성을 상기 제어서버로 전송하는 적어도 하나 이상의 유저모듈;을 포함하는 구성을 개시한다.

대표도

1



(52) CPC특허분류  
*H04N 7/157* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

호스트에게 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저의 음성을 출력하는 호스트모듈;

상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 적어도 하나 이상의 유저모듈에 전송하고, 상기 유저의 음성 데이터를 수신해 상기 호스트모듈에 전송하는 제어서버; 및

상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 상기 유저에게 출력하고, 상기 유저가 입력한 음성을 상기 제어서버로 전송하는 적어도 하나 이상의 유저모듈;을 포함하고,

상기 호스트 모듈은,

상기 가상현실 콘텐츠에 상기 유저의 아바타를 출력하고, 상기 유저의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력하며,

상기 제어서버는,

복수의 상기 유저모듈에서 상기 유저의 음성 데이터를 개별적으로 수신하고,

상기 유저의 음성 데이터를 이용해 유저의 음성을 텍스트로 변환하고, 상기 호스트모듈 및 상기 유저모듈 중 적어도 하나에 전송하고,

상기 호스트모듈 및 상기 유저모듈은 상기 텍스트를 영상에 출력하며,

상기 제어서버는 상기 유저의 음성 데이터를 딥러닝 모델을 통해 발음적 특징을 추출하고, 상기 발음적 특징에 따라 확률이 가장 높은 단어를 1차 단어로 결정하며,

특정 단어에 대한 발음이 유사한 단어들에 대한 유사도를 포함하는 유사군 데이터 및 문장에 있어서 단어들 간에 앞, 뒤로 쓰이는 확률을 학습한 문장연관 데이터를 포함하는 상태에서,

상기 제어서버는 상기 문장연관 데이터에 기초하여 상기 1차 단어의 앞, 뒤 단어들과 연관 확률이 낮은 단어를 수정대상 단어로 결정하고, 상기 수정 대상 단어의 발음 유사군에서 유사도가 특정 확률 이상인 단어들로 상기 1차 단어를 대체하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템.

#### 청구항 2

삭제

#### 청구항 3

삭제

#### 청구항 4

제1항에 있어서,

상기 호스트 모듈은,

출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제하는 것을 특징으로 하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템.

#### 청구항 5

제4항에 있어서,

상기 유저모듈은,

상기 유저의 입력에 따라 상기 텍스트를 수정 및 삭제하고, 수정 및 삭제 결과를 상기 콘텐츠의 영상에 출력하고, 상기 수정 및 삭제 결과를 상기 제어서버로 전송하며,

상기 제어서버는,

상기 수정 및 삭제 결과를 상기 호스트모듈에 전송하며,

상기 호스트모듈은,

상기 수정 및 삭제 결과를 반영해 상기 호스트에게 출력하는 것을 특징으로 하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템.

#### 청구항 6

호스트모듈 및 유저모듈에 제공될 가상현실 콘텐츠가 상기 호스트모듈을 통해 선택되는 단계;

호스트에서 상기 호스트모듈을 이용하여 상기 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저에게 상기 유저모듈을 이용하여 상기 가상현실 콘텐츠의 영상 및 소리를 출력하는 단계; 및

상기 유저가 상기 유저모듈을 통해 입력한 음성을 상기 호스트모듈을 통해 상기 호스트에 출력하는 단계;를 포함하고,

제어서버는, 상기 호스트가 제공하는 상기 가상현실 콘텐츠의 영상 및 소리를 상기 적어도 하나 이상의 유저모듈에 전송하고, 상기 유저의 음성 데이터를 수신해 상기 호스트모듈에 전송하며,

상기 가상현실 콘텐츠의 영상 및 소리를 출력하는 단계는,

상기 가상현실 콘텐츠에 상기 유저의 아바타를 출력하고,

상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계는,

상기 유저의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력하며,

상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계는,

상기 제어 서버가, 복수의 상기 유저의 음성 데이터를 개별적으로 수신하고, 상기 유저의 음성을 텍스트로 변환하고, 상기 텍스트를 영상에 출력하며,

상기 제어 서버는 상기 유저의 음성 데이터를 딥러닝 모델을 통해 발음적 특징을 추출하고, 상기 발음적 특징에 따라 확률이 가장 높은 단어를 1차 단어로 결정하며,

특정 단어에 대한 발음이 유사한 단어들에 대한 유사도를 포함하는 유사군 데이터 및 문장에 있어서 단어들 간에 앞, 뒤로 쓰이는 확률을 학습한 문장연관 데이터를 포함하는 상태에서,

상기 제어서버는 상기 문장연관 데이터에 기초하여 상기 1차 단어의 앞, 뒤 단어들과 연관 확률이 낮은 단어를 수정대상 단어로 결정하고, 상기 수정 대상 단어의 발음 유사군에서 유사도가 특정 확률 이상인 단어들로 상기 1차 단어를 대체하는 것을 특징으로 하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법.

#### 청구항 7

삭제

#### 청구항 8

삭제

**청구항 9**

제6항에 있어서,

상기 유저의 입력에 따라 상기 텍스트를 수정 및 삭제하고, 상기 수정 및 삭제 결과를 상기 콘텐츠의 영상에 출력하는 단계;를 더 포함하는 것을 특징으로 하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법.

**청구항 10**

제9항에 있어서,

출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제하는 단계;를 더 포함하는 것을 특징으로 하는 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법 및 시스템에 있어서, 호스트가 제공하는 영상 또는 콘텐츠가 재생되고 있을 때, 호스트와 유저간에 커뮤니케이션이 가능하고 호스트가 복수의 유저와의 커뮤니케이션을 놓치지 않고 수행할 수 있게 하는 가상공간 내 실시간 양방향 커뮤니케이션 방법 및 시스템에 관한 것이다.

**배경 기술**

[0002] 인간이 가상적으로 만들어내는 가상의 공간을 인공현실(artificial reality), 사이버공간(cyberspace), 가상세계(virtual worlds)라고도 한다. 가장 먼저 가상현실 기법이 적용된 게임의 경우 입체적으로 구성된 화면 속에 게임을 하는 사람이 그 게임의 주인공으로 등장해 문제를 풀어나간다. 이러한 가상현실은 의학 분야에서는 수술 및 해부 연습에 사용되고, 항공·군사 분야에서는 비행조종 훈련에 이용되는 등 각 분야에 도입, 활발히 응용되고 있다.

[0003] 한편, 가상현실(VR·virtual reality)과 현실 세계에 가상정보를 더해 보여주는 기술인 증강현실(AR·augmented reality)을 혼합한 기술은 혼합현실(MR·mixed reality)이라고 한다. VR과 AR, MR은 모두 실제로 존재하지 않은 현실을 구현해 사람이 이를 인지할 수 있도록 하는 기술이라는 점에서 공통점이 있다. 다만 AR은 실제 현실에 가상의 정보를 더해 보여주는 방식이고, VR은 모두 허구의 상황이 제시된다는 점에서 차이가 있다. MR은 AR과 VR을 혼합해 현실 배경에 현실과 가상의 정보를 혼합시켜 제공하는데, 대용량 데이터를 처리할 수 있는 기술이 필요하다.

[0004] HMD(head mounted display)는 VR 체험을 위해 사용자가 머리에 장착하는 디스플레이 디바이스로, 최근에는 사용자의 시각을 외부와 차단한 후 사용자의 시각에 가상세계를 보여줘 실제로 가상세계에 들어와 있는 것처럼 느끼게 하는 역할을 한다. 눈앞에 디스플레이가 오도록 얼굴에 쓰는 형태로 마이크, 스테레오 스피커를 비롯해 여러 센서 등이 탑재돼 있다. VR 헤드셋에 스마트폰을 탑재해 스마트폰 패널을 활용하는 기기는 다이브라고 부른다.

[0005] 이러한 VR 기술이 발전하면서 VR을 활용한 다양한 시도가 이루어지고 있고, 코로나의 창궐로 인해 언택트 기술이 중요한 시대가 도래하면서 VR을 이용한 언택트 세미나, 쇼핑, 교육, 회의 등의 서비스를 제공하기 위한 기술이 필요한 실정이다.

**발명의 내용**

**해결하려는 과제**

[0006] 따라서, 본 발명은 상기한 바와 같은 문제점을 해결하기 위한 것으로서, 가상 공간 내에서 호스트가 제공하는 콘텐츠를 공유하면서 호스트와 유저간에 양방향 커뮤니케이션이 가능한 방법 및 시스템을 제공하고자 한다.

[0007] 본 발명은 가상 공간 호스트와 유저간에 양방향 커뮤니케이션을 수행하면서 복수의 유저가 발생시키는 음성을

놓치지 않고 호스트가 소통할 수 있는 방법 및 시스템을 제공하고자 한다.

**과제의 해결 수단**

- [0008] 상기한 문제를 해결하기 위한 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템은 호스트에게 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저의 음성을 출력하는 호스트모듈; 상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 적어도 하나 이상의 유저모듈에 전송하고, 상기 유저의 음성 데이터를 수신해 상기 호스트모듈에 전송하는 제어서버; 및 상기 호스트가 제공하는 가상현실 콘텐츠의 영상 및 소리를 상기 유저에게 출력하고, 상기 유저가 입력한 음성을 상기 제어서버로 전송하는 적어도 하나 이상의 유저모듈;을 포함할 수 있다.
- [0009] 본 발명의 일 실시 예에 따르면, 상기 호스트 모듈은 상기 가상현실 콘텐츠에 상기 유저의 아바타를 출력하고, 상기 유저의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력할 수 있다.
- [0010] 본 발명의 일 실시 예에 따르면, 상기 제어서버는 복수의 상기 유저모듈에서 상기 유저의 음성 데이터를 개별적으로 수신하고, 상기 유저의 음성 데이터를 이용해 유저의 음성을 텍스트로 변환하고, 상기 호스트모듈 및 상기 유저모듈 중 적어도 하나에 전송하고, 상기 호스트모듈 및 상기 유저모듈은 상기 텍스트를 영상에 출력할 수 있다.
- [0011] 본 발명의 일 실시 예에 따르면, 상기 호스트 모듈은 출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제할 수 있다.
- [0012] 본 발명의 일 실시 예에 따르면, 상기 유저모듈은 상기 유저의 입력에 따라 상기 텍스트를 수정 및 삭제하고, 수정 및 삭제 결과 상기 콘텐츠 영상에 출력하고, 상기 수정 및 삭제 결과를 상기 제어서버로 전송하고, 상기 제어서버는 상기 수정 및 삭제 결과를 상기 호스트모듈에 전송하고 상기 호스트모듈은 상기 수정 및 삭제 결과를 반영해 상기 호스트에게 출력할 수 있다.
- [0013] 상기한 문제를 해결하기 위한 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 호스트가 제공할 가상현실 콘텐츠를 입력하는 단계; 상기 호스트 및 적어도 하나 이상의 유저에게 상기 가상현실 콘텐츠의 영상 및 소리를 출력하는 단계; 및 상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계;를 포함할 수 있다.
- [0014] 본 발명의 일 실시 예에 따르면, 상기 가상현실 콘텐츠의 영상 및 소리를 출력 단계는 상기 가상현실 콘텐츠에 상기 유저의 아바타를 출력하고, 상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계는 상기 유저의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력할 수 있다.
- [0015] 본 발명의 일 실시 예에 따르면, 상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계는 복수의 상기 유저의 음성 데이터를 개별적으로 수신하고, 상기 유저의 음성을 텍스트로 변환하고, 상기 텍스트를 영상에 출력할 수 있다.
- [0016] 본 발명의 일 실시 예에 따르면, 상기 유저의 입력에 따라 상기 텍스트를 수정 및 삭제하고, 상기 수정 및 삭제 결과를 상기 콘텐츠 영상에 출력하는 단계;를 더 포함할 수 있다.
- [0017] 본 발명의 일 실시 예에 따르면, 출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제하는 단계;를 더 포함할 수 있다.

**발명의 효과**

- [0018] 본 발명에 따르면, 가상 공간 내에서 호스트가 제공하는 콘텐츠를 공유하면서 호스트와 유저간에 양방향 커뮤니케이션이 가능하게 할 수 있다.
- [0019] 또한, 복수의 유저와 양방향 커뮤니케이션을 수행하면서도, 호스트가 놓치지 않고 모든 유저의 요구를 충족시킬 수 있는 커뮤니케이션 방법 및 시스템을 제공할 수 있다.
- [0020] 한편, 본 발명의 효과는 이상에서 언급한 효과들로 제한되지 않으며, 이하에서 설명할 내용으로부터 통상의 기술자에게 자명한 범위 내에서 다양한 효과들이 포함될 수 있다.

**도면의 간단한 설명**

- [0021] 도 1은 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템의 설명을 위한 개념도

이다.

도 2는 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템의 블록도이다.

도 3은 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법의 흐름도이다.

도 4는 본 발명의 일 실시 예에 따른 제어서버의 SST 인공지능 학습 네트워크의 일 예시이다.

**발명을 실시하기 위한 구체적인 내용**

- [0022] 이하, 첨부된 도면들을 참조하여 본 발명에 따른 '가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법 및 시스템'을 상세하게 설명한다. 설명하는 실시 예들은 본 발명의 기술사상을 당업자가 용이하게 이해할 수 있도록 제공되는 것으로 이에 의해 본 발명이 한정되지 않는다. 또한, 첨부된 도면에 표현된 사항들은 본 발명의 실시 예들을 쉽게 설명하기 위해 도식화된 도면으로 실제로 구현되는 형태와 상이할 수 있다.
- [0023] 한편, 이하에서 표현되는 각구성부는 본 발명을 구현하기 위한 예일 뿐이다. 따라서, 본 발명의 다른 구현에서는 본 발명의 사상 및 범위를 벗어나지 않는 범위에서 다른 구성부가 사용될 수 있다.
- [0024] 또한, 각구성부는 순전히 하드웨어 또는 소프트웨어의 구성만으로 구현될 수도 있지만, 동일 기능을 수행하는 다양한 하드웨어 및 소프트웨어 구성들의 조합으로 구현될 수도 있다. 또한, 하나의 하드웨어 또는 소프트웨어에 의해 둘 이상의 구성부들이 함께 구현될 수도 있다.
- [0025] 또한, 어떤 구성요소들을 '포함'한다는 표현은, '개방형'의 표현으로서 해당구성요소들이 존재하는 것을 단순히 지칭할 뿐이며, 추가적인 구성요소들을 배제하는 것으로 이해되어서는 안된다.
- [0027] 도 1은 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템의 설명을 위한 개념도이고, 도 2는 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템의 블록도이고, 도 3은 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법의 흐름도이고, 도 4는 본 발명의 일 실시 예에 따른 제어서버의 SST 인공지능 학습 네트워크의 일 예시이다.
- [0028] 도 1 내지 도 4를 참조하면, 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템(1)은 호스트모듈(100), 제어서버(200) 및 유저모듈(300)을 포함할 수 있다. 상기 호스트모듈(100) 또는 상기 유저모듈(300)은 복수의 개별적인 구성으로 포함될 수 있다. 상기 호스트모듈(100) 또는 상기 유저(300) 모듈은 복수의 호스트(10) 또는 복수의 유저(20)를 동일한 가상현실(VR) 콘텐츠에서 실시간으로 양방향 커뮤니케이션을 하도록 할 수 있다.
- [0029] 상기 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템(1)은 호스트(10)와 유저(20)에 콘텐츠를 제공할 수 있다. 상기 콘텐츠는 영상 및 소리를 포함할 수 있다. 상기 콘텐츠는 가상현실(VR) 콘텐츠, 증강현실(AR) 콘텐츠를 포함할 수 있다. 상기 콘텐츠는 상기 호스트(10)가 선택 또는 지정하여 상기 호스트모듈(100) 및 상기 유저모듈(300)을 통해 상기 호스트(10) 및 상기 유저(20)에 제공될 수 있다. 상기 가상현실(VR) 내에서 상기 호스트(10)와 상기 유저(20)는 상호간에 실시간 양방향 커뮤니케이션을 수행할 수 있다. 상기 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템(1)의 호스트모듈(100), 제어서버(200) 및 유저모듈(300)은 네트워크를 이용해 정보를 주고받을 수 있다.
- [0030] 여기서, 네트워크는, 복수의 단말 및 서버들과 같은 각각의 노드 상호 간에 정보 교환이 가능한 연결 구조를 의미하는 것으로, 이러한 네트워크의 일 예에는 RF, 3GPP(3rd Generation Partnership Project) 네트워크, LTE(Long Term Evolution) 네트워크, 5GPP(5rd Generation Partnership Project) 네트워크, WIMAX(World Interoperability for Microwave Access) 네트워크, 인터넷(Internet), LAN(Local Area Network), Wireless LAN(Wireless Local Area Network), WAN(Wide Area Network), PAN(Personal Area Network), 블루투스(Bluetooth) 네트워크, NFC 네트워크, 위성 방송 네트워크, 아날로그 방송 네트워크, DMB(Digital Multimedia Broadcasting) 네트워크 등이 포함되나 이에 한정되지는 않는다.
- [0031] 상기 호스트모듈(100)은 상기 호스트(10)에게 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저(20)의 음성을 출력할 수 있다. 상기 유저(20)의 음성은 상기 유저모듈(300)에서 입력받아 상기 제어서버



(200)를 통해 전송할 수 있다.

- [0032] 상기 호스트모듈(100)은 카메라 및 마이크를 포함할 수 있다. 상기 호스트모듈(100)은 스크린 및 스피커를 포함할 수 있다. 상기 호스트모듈(100)은 상기 콘텐츠의 영상 및 소리를 상기 호스트(10)에게 출력할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)를 촬영하여 상기 호스트(10)의 영상을 상기 제어서버(200)로 전송할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)의 PC, 모바일 디바이스, IoT 디바이스 중 하나를 포함할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)가 착용하는 HMD(head mounted display)일 수 있다. 상기 호스트모듈(100)은 상기 가상현실 콘텐츠 영상에 상기 유저(20)의 아바타를 출력할 수 있다.
- [0033] 상기 호스트모듈(100)은 상기 유저(20)의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력할 수 있다. 예를 들어, 상기 가상현실 콘텐츠에서 복수의 상기 유저(20) 중에서 어느 하나가 상기 유저모듈(300)에 음성을 입력하면 상기 유저모듈(300)은 상기 제어서버(200)로 상기 음성의 데이터를 전송하고, 상기 제어서버(200)는 상기 음성의 데이터를 상기 호스트모듈(100)로 전송하고, 상기 호스트모듈(100)은 상기 음성의 데이터를 소리로 출력할 수 있다. 상기 호스트모듈(100)은 상기 가상현실 콘텐츠 상에 현재 상기 제어서버(200)를 통해 연결된 상기 유저모듈(300)의 수만큼 아바타를 표시할 수 있다. 즉, 상기 아바타는 상기 유저모듈(300)과 1 대 1 대응될 수 있다. 상기 아바타는 상기 제어서버(200)가 생성해 데이터를 상기 호스트모듈(100)로 전송하거나 또는 상기 호스트모듈(100)이 생성하는 것일 수 있다. 상기 아바타는 상기 유저모듈(300)에서도 출력될 수 있다. 상기 호스트모듈(100)은 상기 유저(20)의 음성이 출력될 때 상기 음성을 전송한 상기 유저모듈(300)에 대응되는 아바타에 시각적 효과를 출력할 수 있다. 상기 시각적 효과는 상기 아바타의 밝기 또는 명도 변화, 아이콘, 자막, 깜빡임 등의 효과를 포함할 수 있고, 이에 한정되지 않고 인간이 변화를 인식할 수 있는 모든 시각적 효과를 포함할 수 있다. 상기 호스트모듈(100)은 텍스트를 출력할 수 있다.
- [0034] 상기 제어서버(200)는 상기 호스트모듈(100)과 상기 유저모듈(300)이 주고받는 데이터를 중계할 수 있다. 상기 제어서버(200)는 외부 서버와 연결되어 외부 데이터를 가져와 상기 호스트모듈(100) 및 상기 유저모듈(300)에 전송할 수 있다. 상기 제어서버(200)는 상기 호스트(10)가 선택한 콘텐츠를 상기 호스트모듈(100)을 통해 전송하면 상기 콘텐츠를 상기 유저모듈(300)에 전송할 수 있다. 상기 제어서버(200)는 상기 호스트모듈(100)에서 링크가 전송되면 상기 링크에 포함된 데이터를 수신해 상기 호스트모듈(100) 및 상기 유저모듈(300)에 전송할 수 있다.
- [0035] 상기 제어서버(200)는 상기 호스트(10)가 제공하는 가상현실 콘텐츠의 영상 및 소리를 적어도 하나 이상의 유저모듈(300)에 전송하고, 상기 유저(20)의 음성 데이터를 수신해 상기 호스트모듈(100)에 전송할 수 있다.
- [0036] 상기 제어서버(200)는 복수의 상기 유저모듈(300)에서 상기 유저(20)의 음성 데이터를 개별적으로 수신할 수 있다. 상기 제어서버(200)는 상기 유저(20)의 음성 데이터를 이용해 상기 유저(20)의 음성을 텍스트로 변환하고, 상기 호스트모듈(100) 및 상기 유저모듈(300) 중 적어도 하나에 전송할 수 있다. 상기 호스트모듈(100) 및 상기 유저모듈(300)은 상기 텍스트를 출력할 수 있다.
- [0037] 상기 유저모듈(300)은 상기 호스트(10)가 제공하는 가상현실 콘텐츠의 영상 및 소리를 상기 유저(10)에게 출력하고, 상기 유저(10)가 입력한 음성을 상기 제어서버(200)로 전송할 수 있다.
- [0038] 상기 유저모듈(300)은 상기 유저(10)의 음성을 변환한 텍스트를 상기 유저(20)의 입력에 따라 상기 텍스트를 수정 및 삭제할 수 있다. 상기 유저모듈(300)은 수정 및 삭제 결과 상기 콘텐츠 영상에 출력할 수 있다. 상기 유저모듈(300)은 상기 수정 및 삭제 결과를 상기 제어서버(200)로 전송할 수 있다. 상기 제어서버(200)는 상기 수정 및 삭제 결과를 상기 호스트모듈(100)에 전송할 수 있다. 상기 호스트모듈(100)은 상기 수정 및 삭제 결과를 반영해 상기 호스트에게 출력할 수 있다.
- [0039] 상기 호스트 모듈(100)은 출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제할 수 있다.
- [0040] 본 발명의 일 실시 예에 따르면, 상기 호스트 모듈(100) 또는 상기 유저모듈(300)은 상기 유저(20)의 음성을 변환한 텍스트를 출력할 수 있다. 텍스트가 출력되기 때문에 상기 호스트(10)는 복수의 유저의 문의 사항 등에 대해서 놓치지 않고 대응할 수 있다. 상기 유저(20)는 출력되는 텍스트를 확인하고 상기 호스트(10)가 대응할 필요가 없는 텍스트를 선택해 삭제할 수 있다. 상기 유저(20)는 출력되는 텍스트를 확인하고 상기 텍스트가 상기 유저(20)의 음성을 텍스트로 변환하는 과정에서 오류가 발생한 경우 이를 수정할 수 있다. 상기 호스트(10)는 출력된 텍스트를 확인하고 상기 유저(20)의 요구사항(예를 들면 질문 또는 주문이 될 수 있다.)에 대응할 수 있다. 상기 호스트(10)는 자신이 대응을 마친 텍스트를 선택해 삭제할 수 있다.
- [0041] 상기 제어서버(200)는 STT(Speech-to-Text) 기술을 이용해 상기 유저(20)의 음성 인터페이스를 통해 텍스트(문



자) 데이터를 추출해낼 수 있다.

- [0042] 상기 제어서버(200)는 음향학적 관점에서 말하는 유저, 공간, 노이즈 등의 환경적인 데이터를 이용하고 언어학적 관점에서는 어휘, 문맥, 문법 등을 모델링하기 위한 언어 데이터를 이용해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 음성/언어 데이터로부터 인식 네트워크 모델을 생성하는 오프라인 학습 단계와 사용자가 발성한 음성을 인식하는 온라인 탐색 단계를 통해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 기보유하고 있는 음성과 언어 데이터를 사용해서 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 디코딩 단계에서는 학습 단계 결과인 음향 모델(Acoustic Model), 언어 모델(Language Model)과 발음 사전(Pronunciation Lexicon)을 이용하여 입력된 특징 벡터를 모델과 비교, 스코어링(Scoring)하여 단어 열을 최종 결정할 수 있다.
- [0043] 상기 제어서버(200)는 해당 언어의 음운 환경별 발음의 음향적 특성을 확률 모델로 대표 패턴을 생성하여 음향 모델링을 하고, 어휘 선택, 문장 단위 구문 구조 등 해당 언어의 사용성 문제에 대해 문법 체계를 통째로 학습하여 언어모델링을 할 수 있다. 상기 제어서버(200)는 발음 사전 구축을 위해서는 텍스트를 소리 나는 대로 변환하는 음소 변환(Grapheme-to-Phoneme) 구현을 할 수 있다. 상기 제어서버(200)는 표준 발음을 대상으로 하는 발음 변환 규칙만으로는 방언이나 사용자의 발화 습관과 어투에 따른 다양한 패턴을 반영하기 어려운 경우가 있어 별도의 사전을 구축할 수 있다.
- [0044] 상기 제어서버(200)는 딥러닝(Deep Learning)에 의해 고도화된 음향모델 적응 학습에 기반할 수 있다. 상기 제어서버(200)는 Fully connected DNN(Deep Neural Network), CNN(Convolutional Neural Network)에 기반해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다.
- [0045] 상기 제어서버(200)는 상기 유저(20)의 음성 데이터를 CNN을 통해 분석해 발음적 특징을 추출할 수 있다. 상기 제어서버(200)는 상기 발음적 특징을 추출해 상기 유저(20)의 음성을 단어별로 구간을 분할할 수 있다. 상기 제어서버(200)는 단어 또는 형태소별 발음적 특징을 학습한 데이터를 포함할 수 있다. 상기 제어서버(200)는 단어 또는 형태소별 발음적 특징을 학습한 데이터를 갱신할 수 있다.
- [0046] 상기 제어서버(200)는 상기 단어 또는 형태소의 발음적 특징에 기반해 상기 유저(20)의 음성 데이터에서 분할된 단어를 추정할 수 있다. 상기 제어서버(200)는 기반해 상기 유저(20)의 음성 데이터에서 분할된 단어를 상기 단어 또는 형태소의 발음적 특징에 따라 확률이 가장 높은 단어로 1차적으로 1차 단어로 결정할 수 있다.
- [0047] 상기 제어서버(200)는 특정 단어에 대한 발음이 유사한 단어들과 유사도를 포함하는 발음 유사군 데이터를 포함할 수 있다. 상기 발음 유사군 데이터는 특정 단어가 있으면, 상기 특정 단어와 발음이 유사한 단어들을 유사한 정도에 따라 나열한 데이터를 의미할 수 있다. 상기 제어서버(200)는 문장에 있어서 단어들 간에 앞, 뒤로 쓰이는 확률을 학습한 문장연관 데이터를 포함할 수 있다.
- [0048] 상기 제어서버(200)는 1차적으로 결정한 상기 1차 단어들을 나열할 수 있다. 상기 제어서버(200)는 상기 1차 단어들의 앞, 뒤 단어들과 문장에 함께 쓰일 확률을 상기 문장연관 데이터에 기반해 분석할 수 있다. 상기 제어서버(200)는 상기 문장연관 데이터에 기반해 상기 1차 단어의 앞, 뒤 단어들과 연관 확률이 낮은 단어를 수정 대상 단어로 결정할 수 있다.
- [0049] 상기 제어서버(200)는 상기 수정 대상 단어의 발음 유사군에서 유사도가 특정 확률 이상인 단어들로 대체할 수 있다. 본 발명의 일 실시 예에 따르면, 상기 일정 확률은 80%일 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어를 2차 단어로 결정할 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어가 복수인 경우 연관 확률이 가장 높은 단어를 2차 단어로 결정할 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어가 없는 경우 상기 1차 단어를 2차 단어로 결정할 수 있다. 본 발명의 일 실시 예에 따르면, 상기 임계값은 50%일 수 있다.
- [0050] 상기 제어서버(200)는 상기 연관 확률을 하기 수학적 1에 따라 확률로 연산할 수 있다.
- [0051] [수학적 1]
- [0052] 
$$\sum_{m=1}^n H(S|W_m)$$
- [0053] 여기서, H는 확률함수, S는 단어,  $W_m$ 은 앞, 뒤의 m번째 단어를 의미한다.
- [0054] S에 들어가는 단어는 1차 단어 또는 대체된 단어 중 어느 하나일 수 있다.

- [0055] 본 발명의 일 실시 예에 따라, 확률 또는 임계값을 설정하고 상기 제어서버(200)는 상기 수학식 1에서  $n$ 을 5 이상으로 설정할 수 있다.  $n$ 이 4 이하일 경우 기존의 STT 엔진과 정확도 면에서 큰 차이를 보이지 못하였으나,  $n$ 이 5 이상인 경우 기존의 STT 엔진이 보여주던 오차율이 50%이상 낮아지는 결과를 확인할 수 있었다.
- [0056] 상기 제어서버(200)는 상기 유저(20)가 상기 텍스트를 수정하면, 수정된 결과를 이용해 상기 단어 또는 형태의 발음적 특징을 재학습하여 갱신할 수 있다. 상기 제어서버(200)는 상기 유저(20)가 상기 텍스트를 수정하면, 상기 문장연관 데이터를 재학습해 갱신할 수 있다.
- [0057] 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 호스트가 제공할 가상현실 콘텐츠를 입력하는 단계(S110)를 포함할 수 있다
- [0058] S110 단계에서, 상기 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템(1)은 호스트(10)와 유저(20)에 콘텐츠를 제공할 수 있다. 상기 콘텐츠는 영상 및 소리를 포함할 수 있다. 상기 콘텐츠는 가상현실(VR) 콘텐츠, 증강현실(AR) 콘텐츠를 포함할 수 있다. 상기 콘텐츠는 상기 호스트(10)가 선택 또는 지정하여 상기 호스트모듈(100) 및 상기 유저모듈(300)을 통해 상기 호스트(10) 및 상기 유저(20)에 제공될 수 있다. 상기 가상현실(VR) 내에서 상기 호스트(10)와 상기 유저(20)는 상호간에 실시간 양방향 커뮤니케이션을 수행할 수 있다. 상기 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 시스템(1)의 호스트모듈(100), 제어서버(200) 및 유저모듈(300)은 네트워크를 이용해 정보를 주고받을 수 있다.
- [0059] 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 상기 호스트 및 적어도 하나 이상의 유저에게 상기 가상현실 콘텐츠의 영상 및 소리를 출력하는 단계(S120)를 포함할 수 있다
- [0060] S120 단계에서, 상기 호스트모듈(100)은 카메라 및 마이크를 포함할 수 있다. 상기 호스트모듈(100)은 스크린 및 스피커를 포함할 수 있다. 상기 호스트모듈(100)은 상기 콘텐츠의 영상 및 소리를 상기 호스트(10)에게 출력할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)를 촬영하여 상기 호스트(10)의 영상을 상기 제어서버(200)로 전송할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)의 PC, 모바일 디바이스, IoT 디바이스 중 하나를 포함할 수 있다. 상기 호스트모듈(100)은 상기 호스트(10)가 착용하는 HMD(head mounted display)일 수 있다. 상기 호스트모듈(100)은 상기 가상현실 콘텐츠 영상에 상기 유저(20)의 아바타를 출력할 수 있다.
- [0061] S120 단계에서, 상기 제어서버(200)는 상기 호스트모듈(100)과 상기 유저모듈(300)이 주고받는 데이터를 중계할 수 있다. 상기 제어서버(200)는 외부 서버와 연결되어 외부 데이터를 가져와 상기 호스트모듈(100) 및 상기 유저모듈(300)에 전송할 수 있다. 상기 제어서버(200)는 상기 호스트(10)가 선택한 콘텐츠를 상기 호스트모듈(100)을 통해 전송하면 상기 콘텐츠를 상기 유저모듈(300)에 전송할 수 있다. 상기 제어서버(200)는 상기 호스트모듈(100)에서 링크가 전송되면 상기 링크에 포함된 데이터를 수신해 상기 호스트모듈(100) 및 상기 유저모듈(300)에 전송할 수 있다.
- [0062] 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 상기 유저가 입력한 음성을 상기 호스트에 출력하는 단계(S130)를 포함할 수 있다
- [0063] S130 단계에서, 상기 호스트모듈(100)은 상기 호스트(10)에게 가상현실 콘텐츠의 영상 및 소리를 출력하고, 적어도 하나 이상의 유저(20)의 음성을 출력할 수 있다. 상기 유저(20)의 음성은 상기 유저모듈(300)에서 입력받아 상기 제어서버(200)를 통해 전송할 수 있다.
- [0064] S130 단계에서, 상기 호스트모듈(100)은 상기 유저(20)의 음성이 출력될 때, 음성을 입력한 상기 유저의 아바타에 시각적 효과를 출력할 수 있다. 예를 들어, 상기 가상현실 콘텐츠에서 복수의 상기 유저(20) 중에서 어느 하나가 상기 유저모듈(300)에 음성을 입력하면 상기 유저모듈(300)은 상기 제어서버(200)로 상기 음성의 데이터를 전송하고, 상기 제어서버(200)는 상기 음성의 데이터를 상기 호스트모듈(100)로 전송하고, 상기 호스트모듈(100)은 상기 음성의 데이터를 소리로 출력할 수 있다. 상기 호스트모듈(100)은 상기 가상현실 콘텐츠 상에 현재 상기 제어서버(200)를 통해 연결된 상기 유저모듈(300)의 수만큼 아바타를 표시할 수 있다. 즉, 상기 아바타는 상기 유저모듈(300)과 1 대 1 대응될 수 있다. 상기 아바타는 상기 제어서버(200)가 생성해 데이터를 상기 호스트모듈(100)로 전송하거나 또는 상기 호스트모듈(100)이 생성하는 것일 수 있다. 상기 아바타는 상기 유저모듈(300)에서도 출력될 수 있다. 상기 호스트모듈(100)은 상기 유저(20)의 음성이 출력될 때 상기 음성을 전송한 상기 유저모듈(300)에 대응되는 아바타에 시각적 효과를 출력할 수 있다. 상기 시각적 효과는 상기 아바타의 밝기 또는 명도 변화, 아이콘, 자막, 깜빡임 등의 효과를 포함할 수 있고, 이에 한정되지 않고 인간이 변화를 인식할 수 있는 모든 시각적 효과를 포함할 수 있다. 상기 호스트모듈(100)은 텍스트를 출력할 수 있다.

- [0065] S130 단계에서, 상기 제어서버(200)는 상기 호스트(10)가 제공하는 가상현실 콘텐츠의 영상 및 소리를 적어도 하나 이상의 유저모듈(300)에 전송하고, 상기 유저(20)의 음성 데이터를 수신해 상기 호스트모듈(100)에 전송할 수 있다.
- [0066] S130 단계에서, 상기 제어서버(200)는 복수의 상기 유저모듈(300)에서 상기 유저(20)의 음성 데이터를 개별적으로 수신할 수 있다. 상기 제어서버(200)는 상기 유저(20)의 음성 데이터를 이용해 상기 유저(20)의 음성을 텍스트로 변환하고, 상기 호스트모듈(100) 및 상기 유저모듈(300) 중 적어도 하나에 전송할 수 있다. 상기 호스트모듈(100) 및 상기 유저모듈(300)은 상기 텍스트를 출력할 수 있다.
- [0067] S130 단계에서, 상기 유저모듈(300)은 상기 호스트(10)가 제공하는 가상현실 콘텐츠의 영상 및 소리를 상기 유저(10)에게 출력하고, 상기 유저(10)가 입력한 음성을 상기 제어서버(200)로 전송할 수 있다.
- [0068] S130 단계에서, 본 발명의 일 실시 예에 따르면, 상기 호스트 모듈(100) 또는 상기 유저모듈(300)은 상기 유저(20)의 음성을 변환한 텍스트를 출력할 수 있다. 텍스트가 출력되기 때문에 상기 호스트(10)는 복수의 유저의 문의 사항 등에 대해서 놓치지 않고 대응할 수 있다.
- [0069] S130 단계에서, 상기 제어서버(200)는 STT(Speech-to-Text) 기술을 이용해 상기 유저(20)의 음성 인터페이스를 통해 텍스트(문자) 데이터를 추출해낼 수 있다.
- [0070] S130 단계에서, 상기 제어서버(200)는 음향학적 관점에서 말하는 유저, 공간, 노이즈 등의 환경적인 데이터를 이용하고 언어학적 관점에서는 어휘, 문맥, 문법 등을 모델링하기 위한 언어 데이터를 이용해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 음성/언어 데이터로부터 인식 네트워크 모델을 생성하는 오프라인 학습 단계와 사용자가 발성한 음성을 인식하는 온라인 탐색 단계를 통해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 기보유하고 있는 음성과 언어 데이터를 사용해서 상기 유저(20)의 음성을 텍스트로 변환할 수 있다. 상기 제어서버(200)는 디코딩 단계에서는 학습 단계 결과인 음향 모델(Acoustic Model), 언어 모델(Language Model)과 발음 사전(Pronunciation Lexicon)을 이용하여 입력된 특징 벡터를 모델과 비교, 스코어링(Scoring)하여 단어 열을 최종 결정할 수 있다.
- [0071] S130 단계에서, 상기 제어서버(200)는 해당 언어의 음운 환경별 발음의 음향적 특성을 확률 모델로 대표 패턴을 생성하여 음향 모델링을 하고, 어휘 선택, 문장 단위 구문 구조 등 해당 언어의 사용성 문제에 대해 문법 체계를 통계적으로 학습하여 언어모델링을 할 수 있다. 상기 제어서버(200)는 발음 사전 구축을 위해서는 텍스트를 소리 나는 대로 변환하는 음소 변환(Grapheme-to-Phoneme) 구현을 할 수 있다. 상기 제어서버(200)는 표준 발음을 대상으로 하는 발음 변환 규칙만으로는 방언이나 사용자의 발화 습관과 어투에 따른 다양한 패턴을 반영하기 어려운 경우가 있어 별도의 사전을 구축할 수 있다.
- [0072] S130 단계에서, 상기 제어서버(200)는 딥러닝(Deep Learning)에 의해 고도화된 음향모델 적용 학습에 기반할 수 있다. 상기 제어서버(200)는 Fully connected DNN(Deep Neural Network), CNN(Convolutional Neural Network)에 기반해 상기 유저(20)의 음성을 텍스트로 변환할 수 있다.
- [0073] S130 단계에서, 상기 제어서버(200)는 상기 유저(20)의 음성 데이터를 CNN을 통해 분석해 발음적 특징을 추출할 수 있다. 상기 제어서버(200)는 상기 발음적 특징을 추출해 상기 유저(20)의 음성을 단어별로 구간을 분할할 수 있다. 상기 제어서버(200)는 단어 또는 형태소별 발음적 특징을 학습한 데이터를 포함할 수 있다. 상기 제어서버(200)는 단어 또는 형태소별 발음적 특징을 학습한 데이터를 갱신할 수 있다.
- [0074] S130 단계에서, 상기 제어서버(200)는 상기 단어 또는 형태소의 발음적 특징에 기반해 상기 유저(20)의 음성 데이터에서 분할된 단어를 추정할 수 있다. 상기 제어서버(200)는 기반해 상기 유저(20)의 음성 데이터에서 분할된 단어를 상기 단어 또는 형태소의 발음적 특징에 따라 확률이 가장 높은 단어로 1차적으로 1차 단어로 결정할 수 있다.
- [0075] S130 단계에서, 상기 제어서버(200)는 특정 단어에 대한 발음이 유사한 단어들과 유사도를 포함하는 발음 유사군 데이터를 포함할 수 있다. 상기 발음 유사군 데이터는 특정 단어가 있으면, 상기 특정 단어와 발음이 유사한 단어들을 유사한 정도에 따라 나열한 데이터를 의미할 수 있다. 상기 제어서버(200)는 문장에 있어서 단어들 간에 앞, 뒤로 쓰이는 확률을 학습한 문장연관 데이터를 포함할 수 있다.
- [0076] S130 단계에서, 상기 제어서버(200)는 1차적으로 결정한 상기 1차 단어들을 나열할 수 있다. 상기 제어서버(200)는 상기 1차 단어들의 앞, 뒤 단어들과 문장에 함께 쓰일 확률을 상기 문장연관 데이터에 기반해 분석할 수 있다. 상기 제어서버(200)는 상기 문장연관 데이터에 기반해 상기 1차 단어의 앞, 뒤 단어들과 연관 확률이

낮은 단어를 수정 대상 단어로 결정할 수 있다.

[0077] S130 단계에서, 상기 제어서버(200)는 상기 수정 대상 단어의 발음 유사군에서 유사도가 특정 확률 이상인 단어들로 대체할 수 있다. 본 발명의 일 실시 예에 따르면, 상기 일정 확률은 80%일 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어를 2차 단어로 결정할 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어가 복수인 경우 연관 확률이 가장 높은 단어를 2차 단어로 결정할 수 있다. 상기 제어서버(200)는 대체된 단어 중 연관 확률이 임계 값 이상 높은 단어가 없는 경우 상기 1차 단어를 2차 단어로 결정할 수 있다. 본 발명의 일 실시 예에 따르면, 상기 임계값은 50%일 수 있다.

[0078] S130 단계에서, 상기 제어서버(200)는 상기 연관 확률을 하기 수학적 식 1에 따라 확률로 연산할 수 있다.

[0079] [수학적 식 1]

$$\sum_{m=1}^n H(S|W_m)$$

[0081] 여기서, H는 확률함수, S는 단어,  $W_m$ 은 앞, 뒤의 m번째 단어를 의미한다.

[0082] S에 들어가는 단어는 1차 단어 또는 대체된 단어 중 어느 하나일 수 있다.

[0083] S130 단계에서, 본 발명의 일 실시 예에 따라, 확률 또는 임계값을 설정하고 상기 제어서버(200)는 상기 수학적 식 1에서 n을 5 이상으로 설정할 수 있다. n이 4 이하일 경우 기존의 STT 엔진과 정확도 면에서 큰 차이를 보이지 못하였으나, n이 5 이상인 경우 기존의 STT 엔진이 보여주던 오차율이 50%이상 낮아지는 결과를 확인할 수 있었다.

[0084] S130 단계에서, 상기 제어서버(200)는 상기 유저(20)가 상기 텍스트를 수정하면, 상기 문장연관 데이터를 재학습해 갱신할 수 있다.

[0085] 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 상기 유저의 입력에 따라 상기 텍스트를 수정 및 삭제하고, 상기 수정 및 삭제 결과를 상기 콘텐츠 영상에 출력하는 단계(S140)를 포함할 수 있다

[0086] S140 단계에서, 상기 유저모듈(300)은 상기 유저(10)의 음성을 변환한 텍스트를 상기 유저(20)의 입력에 따라 상기 텍스트를 수정 및 삭제할 수 있다. 상기 유저모듈(300)은 수정 및 삭제 결과 상기 콘텐츠 영상에 출력할 수 있다. 상기 유저모듈(300)은 상기 수정 및 삭제 결과를 상기 제어서버(200)로 전송할 수 있다. 상기 제어서버(200)는 상기 수정 및 삭제 결과를 상기 호스트모듈(100)에 전송할 수 있다. 상기 호스트모듈(100)은 상기 수정 및 삭제 결과를 반영해 상기 호스트에게 출력할 수 있다.

[0087] S140 단계에서, 본 발명의 일 실시 예에 따르면, 상기 유저(20)는 출력되는 텍스트를 확인하고 상기 호스트(10)가 대응할 필요가 없는 텍스트를 선택해 삭제할 수 있다. 상기 유저(20)는 출력되는 텍스트를 확인하고 상기 텍스트가 상기 유저(20)의 음성을 텍스트로 변환하는 과정에서 오류가 발생한 경우 이를 수정할 수 있다.

[0088] 본 발명의 일 실시 예에 따른 가상공간 내 영상 음성 데이터 송수신을 통한 실시간 양방향 커뮤니케이션 방법은 출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제하는 단계(S150)를 포함할 수 있다

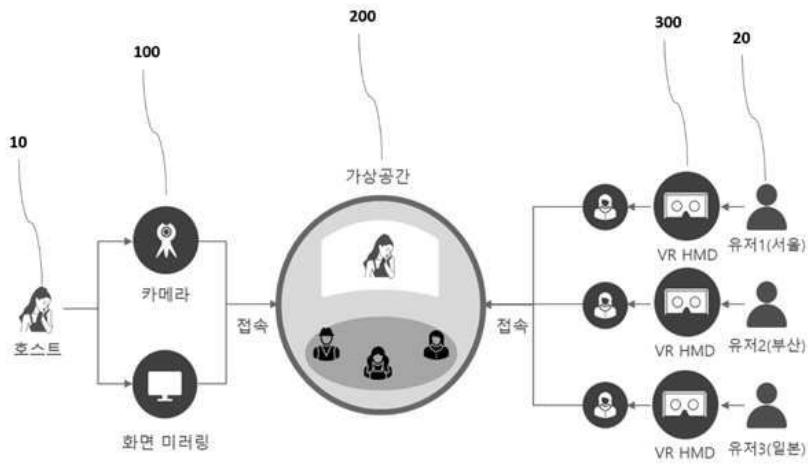
[0089] S150 단계에서, 상기 호스트 모듈(100)은 출력된 복수의 상기 텍스트 중 상기 호스트가 선택한 텍스트를 삭제할 수 있다.

[0090] S150 단계에서, 본 발명의 일 실시 예에 따르면, 상기 호스트(10)는 출력된 텍스트를 확인하고 상기 유저(20)의 요구사항(예를 들면 질문 또는 주문이 될 수 있다.)에 대응할 수 있다. 상기 호스트(10)는 자신이 대응을 마친 텍스트를 선택해 삭제할 수 있다.

[0091] 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

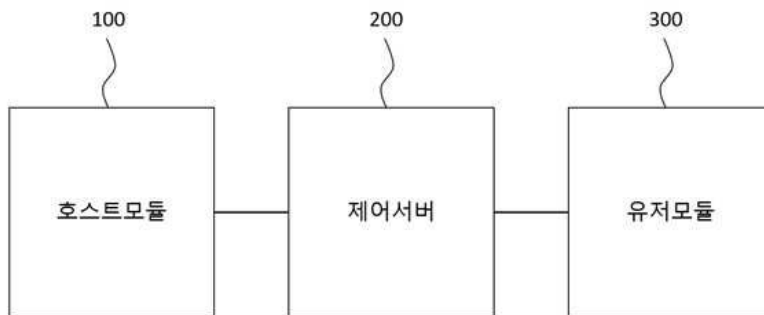
도면

도면1



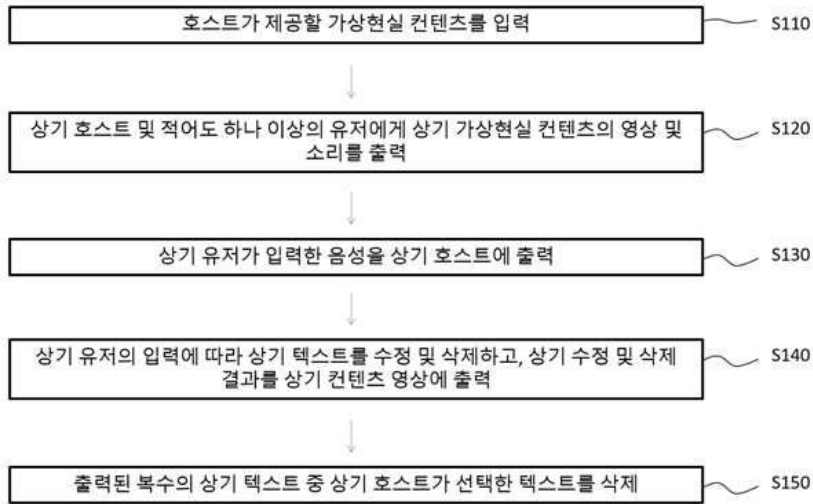
도면2

1





도면3



도면4

