



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0050437
(43) 공개일자 2025년04월15일

(51) 국제특허분류(Int. Cl.)

G06Q 50/50 (2024.01) G06Q 50/10 (2012.01)
G06T 13/40 (2011.01) G06T 19/00 (2011.01)
G10L 13/08 (2006.01) G10L 15/16 (2006.01)
G10L 15/22 (2006.01) G10L 15/26 (2006.01)
G10L 25/51 (2013.01) H04L 51/212 (2022.01)

(52) CPC특허분류

G06Q 50/50 (2024.01)
G06Q 50/10 (2015.01)

(21) 출원번호 10-2023-0133463

(22) 출원일자 2023년10월06일

심사청구일자 없음

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

정문열

서울 동대문구 고산자로32길 78, 청량리역한양수
자인그라시엘 102동 3402호

임영윤

서울특별시 강남구 삼성로4길 17, 504동 606호

김보경

서울특별시 노원구 덕릉로60길 252

(74) 대리인

특허법인 피씨알

전체 청구항 수 : 총 15 항

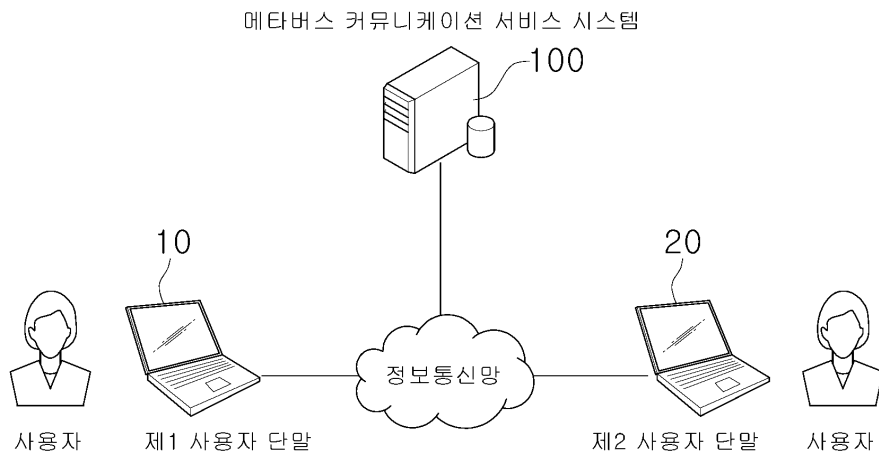
(54) 발명의 명칭 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템 및 방법

(57) 요약

본 발명은 커뮤니케이션 서비스 시스템을 개시한다. 보다 상세하게는, 본 발명은 가상공간 상에서 대화 생성 모델을 이용하여 음성 형태로 화자간 주고 받는 커뮤니케이션 서비스를 제공하는 시스템 및 방법에 관한 것이다.

본 발명의 실시예에 따르면, 메타버스 공간에서 아바타를 통해 사용자가 발화한 음성을 입력받는 대화 생성 모델을 통해 사용자의 연령을 추정하고, 추정된 연령에 적합한 대화 내용을 생성하거나 필터링 함으로써, 아바타와 아바타간 대화에서 언어폭력 발생을 방지하고 원활한 대화가 가능하도록 하며 아바타와 가상객체간 보다 자연스러운 대화가 가능하도록 하는 효과가 있다.

대표도 - 도1



(52) CPC특허분류

G06T 13/40 (2013.01)
G06T 19/006 (2013.01)
G10L 13/08 (2013.01)
G10L 15/16 (2013.01)
G10L 15/22 (2013.01)
G10L 15/26 (2013.01)
G10L 25/51 (2013.01)
H04L 51/212 (2022.05)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179357
과제번호	00156318
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	메타버스 융합대학원(서강대학교)
기 여 율	1/1
과제수행기관명	서강대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

정보통신망을 통해 접속한 하나 이상의 사용자 단말과 데이터 통신을 수행하는 통신부;

하나 이상의 아바타 및 가상객체가 대화를 포함하는 하나 이상의 액티비티를 수행하는 메타버스 공간을 생성하는 메타버스 구현부;

상기 아바타 및 가상객체를 생성하고, 생성된 아바타 및 가상객체를 상기 메타버스 공간상의 임의의 위치에 배치하는 객체 생성부; 및

상기 아바타의 음성을 분석하여 사용자 연령을 추정하고, 추정된 사용자 연령에 따른 대화 내용을 출력하는 대화 생성 모델부를 포함하고,

상기 대화 생성 모델부는,

아바타 및 가상객체간 대화시 사용자 연령에 대응하여 가상객체의 대화 내용을 생성하고, 아바타 및 아바타간 대화시 각 아바타의 사용자 연령에 대응하여 대화 내용의 일부 또는 전부를 변환하는 것인, 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 2

제 1 항에 있어서,

상기 대화 생성 모델부는,

오디오 데이터를 입력받아 이미지 데이터로 변환하는 도메인 변환모듈; 및

상기 이미지 데이터의 특징맵을 추출하는 CNN 및 상기 특징맵에 기초하여 상기 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 분류기를 포함하는 학습 모델

을 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 3

제 2 항에 있어서,

상기 대화 생성 모델부는,

상기 변환모듈을 통해 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터를 입력받아 이미지 데이터로 변환하여 상기 학습 모델을 훈련하는 것인, 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 4

제 2 항 또는 제 3 항에 있어서,

상기 이미지 데이터는,

상기 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram을 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 5

제 2 항에 있어서,

상기 대화 생성 모델부는,

사용자의 음성을 STT API에 입력하여 텍스트로 변환하는 텍스트 변환모듈을 포함하고,

변환된 테스트에 사용자의 분류된 연령대에 따른 연령 레이블을 매핑하는 것인, 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 6

제 5 항에 있어서,

상기 대화 생성 모델부는,

아바타 및 가상객체간 대화인 경우, 상기 텍스트를 대화형 인공지능 플랫폼에 입력하여 상기 연령 레이블을 통해 식별되는 해당 연령대에 대응하는 답변 텍스트를 생성하는 응답모듈;

아바타 및 아바타간 대화인 경우, 연령별 일상대화 데이터셋을 이용하여 학습된 비속어 탐지 모델을 통해 상기 연령 레이블에 따른 사용자 연령별 상기 텍스트에 포함된 비속어를 탐지하여 필터링하는 필터링모듈; 및

상기 답변 텍스트 또는 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하고, 변환된 음성을 상기 가상객체의 답변 또는 아바타간 대화로서 출력하는 대화 출력모듈

을 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 7

제 2 항에 있어서,

제1 언어인 텍스트를 제2 언어로 변환하고, 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하고, 상기 아바타 또는 가상객체의 대화에 따른 음성 출력시 상기 제스처 데이터를 동기화하여 동시에 출력하는 제스처 생성부

를 더 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템.

청구항 8

대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 서비스 방법으로서,

(a) 정보통신망을 통해 하나 이상의 사용자 단말이 접속하면, 사용자의 아바타를 생성하고, 상기 아바타 및 가상객체가 대화를 포함하는 하나 이상의 액티비티를 수행하는 메타버스 공간을 상기 사용자 단말에 제공하는 단계;

(b) 상기 사용자 단말로부터 상기 아바타의 음성을 입력받아 사용자 연령을 추정하는 단계;

(c) 아바타 및 가상객체간 대화시 사용자 연령에 대응하여 가상객체의 대화 내용을 생성하거나, 아바타 및 아바타간 대화시 각 아바타의 사용자 연령에 대응하여 대화 내용의 일부 또는 전부를 변환하는 단계; 및

(d) 추정된 사용자 연령에 따른 대화 내용을 출력하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 9

제 8 항에 있어서,

상기 (b) 단계는,

(b1) 오디오 데이터를 입력받아 이미지 데이터로 변환하는 단계;

(b2) 학습 모델의 CNN을 통해 상기 이미지 데이터의 특징맵을 추출하고, 상기 학습 모델의 분류기를 통해 상기 특징맵에 기초하여 상기 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 10

제 9 항에 있어서,

상기 (b2) 단계 이전에,

(b11) 변환기를 통해 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터를 입력받는 단계; 및

(b12) 상기 오디오 데이터를 이미지 데이터로 변환하여 상기 학습 모델을 훈련하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 11

제 9 항 또는 제 10 항에 있어서,

상기 이미지 데이터는,

상기 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram을 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 12

제 8 항에 있어서,

상기 (b) 단계 및 (c) 단계 사이에,

(b3) 사용자의 음성을 STT API에 입력하여 텍스트로 변환하는 단계; 및

(b4) 변환된 텍스트에 사용자가 속하는 연령 레이블을 매핑하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 13

제 12 항에 있어서,

상기 (c) 단계는, 아바타 및 가상객체간 대화인 경우,

(c1) 상기 연령 레이블을 통해 해당 텍스트의 사용자의 연령을 판단하는 단계;

(c2) 해당 텍스트를 대화형 인공지능 플랫폼에 입력하여 해당 연령대에 대응하는 답변 텍스트를 생성하는 단계;

(c3) 상기 답변 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 상기 가상객체의 답변으로서 생성하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 14

제 12 항에 있어서,

상기 (c) 단계는, 아바타 및 아바타간 대화인 경우,

(c4) 상기 연령 레이블을 통해 해당 텍스트의 사용자의 연령을 판단하는 단계;

(c5) 연령별 일상대화 데이터셋을 이용하여 학습된 비속어 탐지 모델을 통해 상기 연령 레이블에 따른 사용자 연령별 상기 텍스트에 포함된 비속어를 탐지하여 필터링하는 단계; 및

(c6) 상기 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 아바타간 대화로서 생성하는 단계

를 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

청구항 15

제 8 항에 있어서,

상기 단계 (d)는,

(d1) 제1 언어인 텍스트를 제2 언어로 변환하는 단계;

(d2) 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하는 단계; 및

(d3) 상기 아바타 또는 가상객체의 대화에 따른 음성 출력시 상기 제스처 데이터를 동기화하여 동시에 출력하는 단계

를 더 포함하는 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법.

발명의 설명

기술 분야

[0001] 본 발명은 커뮤니케이션 서비스 시스템에 관한 것으로, 특히 가상공간 상에서 대화 생성 모델을 이용하여 음성 형태로 화자간 주고 받는 커뮤니케이션 서비스를 제공하는 시스템 및 방법에 관한 것이다.

배경 기술

[0002] 메타버스란, 현실세계와 같은 사회, 경제, 문화 활동이 이뤄지는 3차원 가상세계를 가리키는 용어로서, 메타버스는 '가상', '초월' 등을 뜻하는 영어 단어 '메타'(Meta)와 우주를 뜻하는 '유니버스'(Universe)의 합성어이다.

[0003] 특히, 메타버스 공간에서는 VR, 컴퓨터로 만들어 놓은 가상의 세계에서 사람이 실제와 같은 체험을 할 수 있도록 하는 최첨단 기술이 적용되어 사용자를 대체하는 아바타라는 개체를 통해 단지 게임이나 가상현실을 즐기는 데 그치지 않고 실제 현실공간에서 수행하는 사회, 문화적 활동을 할 수 있다는 특징이 있다.

[0004] 최근, 이러한 메타버스에 대한 수요가 증가함에 따라, 가상공간에서 아바타와 게임, 대화 등의 액티비티를 함께 수행할 수 있는 가상인간을 포함하는 가상객체가 등장하고 있으며, 이러한 가상객체는 실제 인간들의 물리적인 한계를 보완할 수 있어 부가가치가 크고 시장에서의 성장가능성도 무한하다고 평가받고 있다.

[0005] 특히, 기업들은 가상객체를 영구적이고 영속적인 비즈니스 파트너로 활용 가능하다고 판단하고 있으며, 이에 메타버스를 기반으로 하는 다양한 비즈니스 모델을 연구하고 있다.

[0006] 그러나, 메타버스에서 가상객체와 사용자의 아바타가 원활하게 상호작용을 하기 위해서는 텍스트 중심의 인터페이스 뿐만 아니라, 음성 대화와 같은 사실적인 경험을 제공할 수 있는 멀티모달(Multimodal) 대화 생성 시스템의 구축이 요구되고 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 등록특허공보 제10-2383574호(공고일자: 2022.04.08.)

(특허문헌 0002) 공개특허공보 제10-2023-0112390(공개일자: 2023.07.27.)

발명의 내용

해결하려는 과제

[0008] 본 발명은 기존의 텍스트 기반 대화 생성 시스템에서 진보된 음성 데이터 기반 커뮤니케이션 시스템을 제공하되, 아바타 및 아바타 뿐만 아니라, 아바타 및 가상객체간 연령별로 맞춤형 대화가 가능한 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템 및 방법을 제공하는 데 과제가 있다.

과제의 해결 수단

[0009] 진술한 과제를 해결하기 위해, 본 발명의 일 실시예에 따른 화 생성 모델 기반 메타버스 커뮤니케이션 서비스 방법은, 정보통신망을 통해 접속한 하나 이상의 사용자 단말과 데이터 통신을 수행하는 통신부, 하나 이상의 아바타 및 가상객체가 대화를 포함하는 하나 이상의 액티비티를 수행하는 메타버스 공간을 생성하는 메타버스 구현부, 상기 아바타 및 가상객체를 생성하고, 생성된 아바타 및 가상객체를 상기 메타버스 공간상의 임의의 위치에 배치하는 객체 생성부 및, 상기 아바타의 음성을 분석하여 사용자 연령을 추정하고, 추정된 사용자 연령에

다른 대화 내용을 출력하는 대화 생성 모델부를 포함하고, 상기 대화 생성 모델부는, 아바타 및 가상객체간 대화시 사용자 연령에 대응하여 가상객체의 대화 내용을 생성하고, 아바타 및 아바타간 대화시 각 아바타의 사용자 연령에 대응하여 대화 내용의 일부 또는 전부를 변환하는 것일 수 있다.

- [0010] 상기 대화 생성 모델부는, 오디오 데이터를 입력받아 이미지 데이터로 변환하는 도메인 변환모듈 및, 상기 이미지 데이터의 특징맵을 추출하는 CNN 및 상기 특징맵에 기초하여 상기 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 분류기를 포함하는 학습 모델을 포함할 수 있다.
- [0011] 상기 대화 생성 모델부는, 상기 변환모듈을 통해 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터를 입력받아 이미지 데이터로 변환하여 상기 학습 모델을 훈련하는 것일 수 있다.
- [0012] 상기 이미지 데이터는, 상기 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram을 포함할 수 있다.
- [0013] 상기 대화 생성 모델부는, 사용자의 음성을 STT API에 입력하여 텍스트로 변환하는 텍스트 변환모듈을 포함하고, 변환된 텍스트에 사용자의 분류된 연령대에 따른 연령 레이블을 매핑할 수 있다.
- [0014] 상기 대화 생성 모델부는, 아바타 및 가상객체간 대화인 경우, 상기 텍스트를 대화형 인공지능 플랫폼에 입력하여 상기 연령 레이블을 통해 식별되는 해당 연령대에 대응하는 답변 텍스트를 생성하는 응답모듈, 아바타 및 아바타간 대화인 경우, 연령별 일상대화 데이터셋을 이용하여 학습된 비속어 탐지 모델을 통해 상기 연령 레이블에 따른 사용자 연령별 상기 텍스트에 포함된 비속어를 탐지하여 필터링하는 필터링모듈 및, 상기 답변 텍스트 또는 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하고, 변환된 음성을 상기 가상객체의 답변 또는 아바타간 대화로서 출력하는 대화 출력모듈을 포함할 수 있다.
- [0015] 상기 시스템은, 제1 언어인 텍스트를 제2 언어로 변환하고, 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하고, 상기 아바타 또는 가상객체의 대화에 따른 음성 출력시 상기 제스처 데이터를 동기화하여 동시에 출력하는 제스처 생성부를 더 포함할 수 있다.
- [0016] 또한, 본 발명의 일 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 서비스 방법은, (a) 정보통신망을 통해 하나 이상의 사용자 단말이 접속하면, 사용자의 아바타를 생성하고, 상기 아바타 및 가상객체가 대화를 포함하는 하나 이상의 액티비티를 수행하는 메타버스 공간을 상기 사용자 단말에 제공하는 단계, (b) 상기 사용자 단말로부터 상기 아바타의 음성을 입력받아 사용자 연령을 추정하는 단계, (c) 아바타 및 가상객체간 대화시 사용자 연령에 대응하여 가상객체의 대화 내용을 생성하거나, 아바타 및 아바타간 대화시 각 아바타의 사용자 연령에 대응하여 대화 내용의 일부 또는 전부를 변환하는 단계 및, (d) 추정된 사용자 연령에 따른 대화 내용을 출력하는 단계를 포함할 수 있다.
- [0017] 상기 (b) 단계는, (b1) 오디오 데이터를 입력받아 이미지 데이터로 변환하는 단계, (b2) 학습 모델의 CNN을 통해 상기 이미지 데이터의 특징맵을 추출하고, 상기 학습 모델의 분류기를 통해 상기 특징맵에 기초하여 상기 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 단계를 포함할 수 있다.
- [0018] 상기 (b2) 단계 이전에, (b11) 변환기를 통해 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터를 입력받는 단계 및, (b12) 상기 오디오 데이터를 이미지 데이터로 변환하여 상기 학습 모델을 훈련하는 단계를 포함할 수 있다.
- [0019] 상기 이미지 데이터는, 상기 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram을 포함할 수 있다.
- [0020] 상기 (b) 단계 및 (c) 단계 사이에, (b3) 사용자의 음성을 STT API에 입력하여 텍스트로 변환하는 단계 및, (b4) 변환된 텍스트에 사용자가 속하는 연령 레이블을 매핑하는 단계를 포함할 수 있다.
- [0021] 상기 (c) 단계는, 아바타 및 가상객체간 대화인 경우, (c1) 상기 연령 레이블을 통해 해당 텍스트의 사용자의 연령을 판단하는 단계, (c2) 해당 텍스트를 대화형 인공지능 플랫폼에 입력하여 해당 연령대에 대응하는 답변 텍스트를 생성하는 단계 및, (c3) 상기 답변 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 상기 가상객체의 답변으로서 생성하는 단계를 포함할 수 있다.
- [0022] 상기 (c) 단계는, 아바타 및 아바타간 대화인 경우, (c4) 상기 연령 레이블을 통해 해당 텍스트의 사용자의 연령을 판단하는 단계, (c5) 연령별 일상대화 데이터셋을 이용하여 학습된 비속어 탐지 모델을 통해 상기 연령 레

이블에 따른 사용자 연령별 상기 텍스트에 포함된 비속어를 탐지하여 필터링하는 단계 및, (c6) 상기 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 아바타간 대화로서 생성하는 단계를 포함할 수 있다.

- [0023] 상기 단계 (d)는, (d1) 제1 언어인 텍스트를 제2 언어로 변환하는 단계, (d2) 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하는 단계 및, (d3) 상기 아바타 또는 가상객체의 대화에 따른 음성 출력시 상기 제스처 데이터를 동기화하여 동시에 출력하는 단계를 더 포함할 수 있다.

발명의 효과

- [0024] 본 발명의 실시예에 따르면, 메타버스 공간에서 아바타를 통해 사용자가 발화한 음성을 입력받는 대화 생성 모델을 통해 사용자의 연령을 추정하고, 추정된 연령에 적합한 대화 내용을 생성하거나 필터링 함으로써, 아바타와 아바타간 대화에서 언어폭력 발생을 방지하고 원활한 대화가 가능하도록 하며 아바타와 가상객체간 보다 자연스러운 대화가 가능하도록 하는 효과가 있다.

- [0025] 또한, 본 발명의 실시예에 따르면, 음성 특성에 따라 추정된 사용자의 연령에 대응하여 아바타의 외관을 결정하거나, 대화 내용에 대응하여 아바타가 취하는 동작을 제어함으로써, 메타버스 공간내에서 보다 대화자간 보다 자연스럽고 역동적인 커뮤니케이션을 구현할 수 있는 효과가 있다.

도면의 간단한 설명

- [0026] 도 1은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 전체 구조를 개략적으로 나타낸 도면이다.

도 2는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 구조를 나타낸 도면이다.

도 3은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 대화 생성 모델부의 구조를 나타낸 도면이다.

도 4는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 적용된 학습 모델을 나타낸 도면이다.

도 5는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에서 대화형 인공지능 플랫폼을 통한 대화 생성 형태를 예시한 도면이다.

도 6은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 적용된 제스처 생성 기법을 예시한 도면이다.

도 7은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 커뮤니케이션 서비스 제공방법을 나타낸 도면이다.

도 8 및 도 9는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 연령 추정 정확도 및 대화 생성 모델에 의한 생성 결과를 예시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 설명에 앞서, 명세서 전체에서 어떤 부분이 어떤 구성요소를 "구비" 또는 "포함" 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한, 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 전체한다. 또한, 명세서에 기재된 "...부(Unit)", "...서버(Server)", "...시스템(System)" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어, 소프트웨어 또는, 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

- [0028] 또한, 본 명세서에서 "실시예"라는 용어는 예시, 사례 또는 도해의 역할을 하는 것을 의미하나, 발명의 대상은 그러한 예에 의해 제한되지 않는다. 또한, "포함하는", "구비하는", "갖는" 및 다른 유사한 용어가 사용되고 있으나, 청구범위에서 사용되는 경우 임의의 추가적인 또는 다른 구성요소를 배제하지 않는 개방적인 전환어(Transition word)로서 "포함하는(Comprising)"이라는 용어와 유사한 방식으로 포괄적으로 사용된다.

- [0029] 본 명세서에 설명된 다양한 기법은 하드웨어 또는 소프트웨어와 함께 구현될 수 있거나, 적합한 경우에 이들 모두의 조합과 함께 구현될 수 있다. 본 명세서에 사용된 바와 같은 "...부(Unit)", "...서버(Server)" 및 "...시

스텝(System)" 등의 용어는 마찬가지로 컴퓨터 관련 엔티티(Entity), 즉 하드웨어, 하드웨어 및 소프트웨어의 조합, 소프트웨어 또는 실행 시의 소프트웨어와 등가로 취급할 수 있다. 또한, 본 발명의 시스템에서 구현하는 각 기능은 모듈단위의 프로그램으로 구성될 수 있고, 하나의 물리적 메모리에 기록되거나, 둘 이상의 메모리 및 기록매체 사이에 분산되어 기록될 수 있다.

- [0030] 이하의 설명에서, 본 발명의 "대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템"을 가리키는 용어는, "커뮤니케이션 서비스 시스템" 또는 "시스템"으로 혼용 기재될 수 있다.
- [0031] 이하, 도면을 참조하여 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템을 상세히 설명한다.
- [0032] 도 1은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 전체 구조를 개략적으로 나타낸 도면이다.
- [0033] 도 1을 참조하면, 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템(100)은, 하나의 사용자 단말(10)의 요청에 따라 사용자와 가상공간내 존재하는 가상객체간의 대화, 또는 제1 사용자 단말(10)과 제2 사용자 단말(20)간의 가상공간내에서의 대화 서비스를 제공할 수 있다.
- [0034] 사용자 단말(10, 20)은 본 발명의 시스템(100)이 제공하는 커뮤니케이션 서비스를 이용하고자 하는 사용자들이 소지, 사용하는 하는 단말 장치로서, 전용 어플리케이션 또는 웹 브라우저를 실행하여 시스템(100)에 접속 및 서비스를 이용할 수 있도록 한다.
- [0035] 전술한 사용자 단말(10, 20)로는, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet) 단말, 스마트폰(smartphone), 스마트패드(smartpad), 태블릿 PC(Tablet PC) 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치 및 태블릿PC, 노트북과 같은 컴퓨팅 장치가 이용될 수 있으나, 이에 한정되는 것은 아니다.
- [0036] 특히, 본 발명의 커뮤니케이션 서비스는 가상공간내에서 사용자와 가상객체, 또는 사용자와 타사용자간의 채팅과 같은 대화 서비스를 가리키는 것으로, 사용자 단말(10, 20)에 탑재되는 화면을 통해 기존의 텍스트(Text) 기반의 대화가 아닌, 가상공간 상에서 사용자를 대체하는 아바타와 NPC 형태의 가상객체 또는 타 사용자의 아바타간 1인칭 또는 3인칭 시점의 사운드(Sound) 및 비주얼(Visual) 기반의 대화를 수행하는 것을 특징으로 한다.
- [0037] 커뮤니케이션 서비스 시스템(100)은, 정보통신망을 통해 연결된 다수의 사용자 단말(10, 20)의 요청에 따라 메타버스 기반 가상공간과, 그 가상공간에서 아바타 및 가상객체간 다양한 상호작용을 제공하는 서버 장치로서, 특히 아바타에 의한 커뮤니케이션 서비스를 제공할 수 있다.
- [0038] 여기서, 시스템(100)과 사용자 단말(10, 20)간 연결을 위한 정보통신망으로는, 공지의 데이터 통신망으로서 LAN(Local Area Network), WAN(Wide Area Network)등의 폐쇄형 네트워크를 비롯하여 인터넷과 같은 개방형 네트워크가 이용될 수 있다. 인터넷은 TCP/IP 프로토콜 및 그 상위계층에 존재하는 여러 서비스, 즉 HTTP(HyperText Transfer Protocol), Telnet, FTP(File Transfer Protocol), DNS(Domain Name System), SMTP(Simple Mail Transfer Protocol), SNMP(Simple Network Management Protocol), NFS(Network File Service), NIS(Network Information Service)를 제공하는 전세계적인 개방형 컴퓨터 네트워크 구조를 의미할 수 있다. 또한, 무선 통신망의 일 예로는, 이동통신을 위한 기술표준들 또는 통신방식(예를 들어, GSM(Global System for Mobile communication), 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), 5G 등 중, 어느 하나가 이용될 수 있으나 이에 한정되는 것은 아니다.
- [0039] 특히, 본 발명의 실시예에 따르면, 시스템(100)은 아바타 및 가상객체간 대화 또는 두 아바타간의 대화에 있어서, 사운드 기반 대화를 수행함에 따라 실시간으로 화자의 음성을 분석하여 연령을 추정하고, 추정된 연령에 맞추어 대화 내용의 일부 또는 전부를 변환하는 것을 특징으로 한다.
- [0040] 이는 다양한 연령층이 제한없이 접속할 수 있는 메타버스 환경에서 저연령 사용자를 보호하고, 상호간 발생할 수 있는 언어/사이버 폭력을 방지하기 위함이다.
- [0041] 이를 위해, 시스템(100)은 공개된 다양한 연령층(5세 ~ 89세)에 대한 오디오 학습 데이터를 이용하여 훈련된 학

습 모델을 준비하고, 사용자 단말(10, 20)을 통해 실시간으로 입력되는 음성을 이미지 포맷 데이터로 변환하여 학습 모델을 통해 연령을 추정하고 특정 연령대 그룹으로 분류할 수 있다.

[0042] 또한, 시스템(100)은 현재 발화자의 음성을 텍스트화하고 대화 생성 모델을 통해 내용을 분석하여 문장을 연령에 적합한 형태로 단어를 변경하거나 문장 구조를 변환하고, 그 변경된 텍스트를 다시 음성 형태로 변환하여 출력함으로써 실시간 대화가 이루어지도록 할 수 있다.

[0043] 즉, 본 발명의 커뮤니케이션 서비스 시스템(100)은 아바타-아바타간, 아바타-가상객체간 실시간 대화에서 화자의 연령을 추정하고 대화 생성 모델을 통해 연령에 맞게 대화 내용을 변환함으로써, 온라인 상의 대화에서 발생할 수 있는 사고를 최소화할 수 있다. 이를 위한 시스템(100)에 관한 상세한 설명은 후술한다.

[0044] 전술한 구조에 따라, 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템은, 다음과 같은 효과를 제공할 수 있을 것으로 기대된다.

[0045] 첫째, '사용자 경험의 향상'으로서, 메타버스 공간에서 사용자는 더욱 몰입감 있고 자유로운 움직임을 경험할 수 있음에 따라, 사용자들이 가상공간에서 보다 현실적이고 자연스러운 상호 작용을 할 수 있게 된다.

[0046] 둘째, '참여와 상호작용의 증진'으로서, 메타버스에서의 참여와 상호작용이 증진되며, 사용자는 연령에 따라 자동적으로 조정되는 제스처를 통해 가상 환경에 직접적으로 개입할 수 있어 참여도와 참여감이 높아지게 된다. 또한, 다른 사용자와의 상호작용도 활발해짐에 따라, 제스처를 통해 소통하고 협업하는 경험을 얻을 수 있다.

[0047] 셋째, '창의적인 표현과 활동 강화'로서, 사용자의 창의적인 표현과 다양한 활동 영역을 지원하게 된다. 사용자는 독특한 동작이나 제스처를 통해 자신만의 개성을 가진 캐릭터를 만들고, 메타버스에서 움직임과 행동을 표현할 수 있게 된다. 이는 창의성을 자극하고 새로운 경험을 제공함으로써 사용자의 참여와 만족도를 높여줄 수 있다.

[0048] 넷째, '접근성과 포용성 강화'로서, 연령을 고려한 대화 생성 모델을 사용함에 따라 언어나 기술적인 장벽을 넘어서는 접근성과 포용성을 강화할 수 있으며, 이를 통해 의사소통에 어려움을 겪는 사람들이 제스처를 통해 메타버스에 참여하고 표현할 수 있도록 한다. 이를 통해 차별화되고 포용적인 가상공간을 조성하는데 기여할 것으로 기대된다.

[0049] 이하, 도면을 참조하여 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템을 보다 상세히 설명한다.

[0050] 도 2는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 구조를 나타낸 도면이다.

[0051] 도 2를 참조하면, 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템(100)은, 정보통신망을 통해 접속한 하나 이상의 사용자 단말과 데이터 통신을 수행하는 통신부(110), 하나 이상의 아바타 및 가상객체가 대화를 포함하는 하나 이상의 액티비티를 수행하는 메타버스 공간을 생성하는 메타버스 구현부(120), 아바타 및 가상객체를 생성하고, 생성된 아바타 및 가상객체를 메타버스 공간상의 임의의 위치에 배치하는 객체 생성부(130), 아바타의 음성을 분석하여 사용자 연령을 추정하고, 추정된 사용자 연령에 따른 대화 내용을 출력하는 대화 생성 모델부(200), 제1 언어인 텍스트를 제2 언어로 변환하고, 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하고, 아바타 또는 가상객체의 대화에 따른 음성 출력시 제스처 데이터를 동기화하여 동시에 출력하는 제스처 생성부(300)를 포함할 수 있다.

[0052] 특히, 본 발명의 실시예에 따르면, 상기 대화 생성 모델부(200)는, 아바타 및 가상객체간 대화시 사용자 연령에 대응하여 가상객체의 대화 내용을 생성하고, 아바타 및 아바타간 대화시 각 아바타의 사용자 연령에 대응하여 대화 내용의 일부 또는 전부를 변환하는 것을 특징으로 한다.

[0053] 상세하게는, 통신부(110)는 정보통신망을 통해 접속한 사용자 단말(10, 20)과 데이터 통신을 수행하는 것으로, 메타버스 공간을 사용자 단말(10, 20)에 제공하며, 음성 데이터를 입력받아 사용자 단말(10, 20) 각각과 가상객체와의 대화 또는 제1 사용자 단말(10)과 제2 사용자 단말(20)간 대화에 따른 커뮤니케이션 서비스에 따른 오디오, 영상 등을 제공할 수 있다.

[0054] 메타버스 구현부(120)는 사용자에게 대응하는 다수의 아바타 및 가상객체가 서로 상호작용을 수행할 수 있는 메타버스 가상 공간을 구현할 수 있다. 이러한 메타버스 공간은 가상의 여행, 쇼핑 및 게임 등 여러 형태의 액티비티를 수행할 수 있는 공간이며, 아바타들과 가상객체들이 음성에 기반한 커뮤니케이션 서비스를 이용할 수 있는

공간일 수 있다.

- [0055] 객체 생성부(130)는 사용자에게 대응하는 아바타를 비롯하여 생성된 가상 공간내에서 활동하는 주체인 다양한 가상 객체(virtual object)를 생성하고, 이를 메타버스 공간에 적용할 수 있다.
- [0056] 이러한 가상 객체는 설정된 로직에 따라 자동으로 소정의 행위를 수행하는 NPC(Non Player Character)를 비롯하여, 가상의 배경, 건물, 자연환경 등의 가상 공간상에서 구현될 수 있는 다양한 객체를 가리킬 수 있다.
- [0057] 대화 생성 모델부(200)는 본 발명의 커뮤니케이션 서비스를 구현하기 위한 주요 구성부로서, 사용자 단말(10, 20)로부터 입력되는 사용자의 음성을 분석하여 연령을 추정하고, 추정된 연령에 따라 사용자를 비롯한 가상 객체의 음성 내용을 필터링하거나 변환하고, 응답을 생성하여 음성 데이터로 출력할 수 있다.
- [0058] 이를 위해, 대화 생성 모델부(200)는 외부의 STT API 및 TTS API를 비롯하여 대화형 인공지능 플랫폼과 연동할 수 있고, 오디오 데이터의 학습을 통해 음성에 따른 연령을 추정하는 학습 모델을 포함할 수 있으며, 이를 통해 아바타 및 아바타간 대화에서 비속어를 필터링하거나, 사용자의 연령에 적합한 형태로 대화 문장을 변경할 수 있다.
- [0059] 전술한 기능을 구현하기 위한 대화 생성 모델부(200)에 대한 상세한 설명은 후술한다.
- [0060] 또한, 본 발명의 일 실시예에 따르면, 대화중인 아바타 및 가상 객체는 대화 내용에 따라 적합한 표정, 손짓 등의 제스처를 취할 수 있으며, 이를 위한 제스처 생성부(300)를 더 포함할 수 있다.
- [0061] 이를 위해, 제스처 생성부(300)는 사용자로부터 입력되는 한국어 음성 데이터를 텍스트 데이터로 입력받아 영어 텍스트 데이터로 변환할 수 있고, 그 변환된 텍스트에 기반하여 제스처 텍스트를 생성하는 Emotion Classification 입력할 수 있다. 그리고, 제스처 생성부(300)는 입력된 제스처 텍스트를 해당 아바타 또는 가상 객체에 동기화하여 음성과 함께 출력되도록 함으로써, 메타버스 공간에서 보다 역동적인 커뮤니케이션 서비스를 구현할 수 있도록 한다.
- [0062] 예를 들면, 제스처 생성부(300)는 공지의 오픈소스, 일예로 KoBERT를 활용하여 제작된 현재 대화의 주제, 화행을 분포를 분류하는 모델(Conversational Categorization Model)이 적용될 수 있으며, 아바타 및 가상객체의 발화(text)를 상기 모델에 입력하여 80개의 제스처 중 1개를 아바타 또는 가상객체에 리타겟팅(retargeting)할 수 있다.
- [0063] 여기서, 리타겟팅은 기본적인 사람 형태를 갖는 캐릭터가 공통적으로 하나의 머리, 척추, 골반, 두 팔 및 두 다리로 이루어짐에 따라, 이러한 공통적인 구조를 이용해서 애니메이션을 재사용할 수 있도록 만든 기술을 가리킨다.
- [0064] 이하, 도면을 참조하여 전술한 대화 생성 모델부의 구조에 대하여 구체적으로 설명한다.
- [0065] 도 3은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템의 대화 생성 모델부의 구조를 나타낸 도면이다.
- [0066] 도 3을 참조하면, 본 발명의 실시예에 따른 대화 생성 모델부(200)는 오디오 데이터를 입력받아 이미지 데이터로 변환하는 도메인 변환모듈(210), 이미지 데이터의 특징맵을 추출하는 CNN(221) 및, 특징맵에 기초하여 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 분류기(223)를 포함하는 학습 모델(220), 사용자의 음성을 STT API에 입력하여 텍스트로 변환하는 텍스트 변환모듈(230), 아바타 및 가상객체간 대화인 경우, 텍스트에 연령 레이블을 결합하고, 대화형 인공지능 플랫폼에 입력하여 해당 연령대에 대응하는 답변 텍스트를 생성하는 응답 모듈(250), 아바타 및 아바타간 대화인 경우, 연령별 일상대화 데이터셋을 이용하여 학습된 비속어 탐지 모델을 통해 연령 레이블에 따른 사용자 연령별 텍스트에 포함된 비속어를 탐지하여 필터링하는 필터링 모듈(260) 및, 답변 텍스트 또는 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하고, 변환된 음성을 가상객체의 답변 또는 아바타간 대화로서 출력하는 대화 출력 모듈(270)을 포함할 수 있다.
- [0067] 도메인 변환모듈(210)은 오디오 데이터(wav)를 입력받아, 오디오 도메인을 이미지 도메인의 데이터로 변환할 수 있다. 이때, 이미지 데이터(img)는 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram 일 수 있다.
- [0068] 즉, 도메인 변환모듈(210)은 오디오 데이터를 Mel-Spectrogram로 변환하여 후술하는 학습 모델(220)에 입력함으로써 학습 모델을 훈련하거나, 오디오 데이터에 대한 연령을 추정 및 분류하도록 한다.

- [0069] 특히, 본 발명의 실시예에 따른 도메인 변환모듈(210)은 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터(wav)를 입력받아 이미지 데이터(img)로 변환할 수 있고, 학습 모델(220)은 이미지 데이터(img)를 이용해 훈련하는 것을 특징으로 한다.
- [0070] 학습 모델(220)은 도메인 변환모듈(210)로부터 제공되는 이미지 데이터(img)를 통해 기계학습을 수행하고 오디오 데이터에 대한 연령을 추정 및 분류할 수 있다. 이를 위해 학습 모델(220)은 컨볼루션 신경망 네트워크(CNN) 및 분류기(Linear Classifier)를 포함할 수 있다.
- [0071] 도 4는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 적용된 학습 모델을 나타낸 도면으로서, 도 4를 함께 참조하면, 대화 생성 모델부(200)에 입력되는 오디오 데이터(Audio wave)는 도메인 변환모듈(210)에 의해 이미지 데이터(img)인 Mel-Spectrogram으로 변환되고, Mel-Spectrogram은 CNN Architecture를 통해 학습되어 특징 맵(Feature Map)을 도출하는 데 이용된다. 이러한 특징 맵(Feature Map)을 통해 오디오의 특성이 추정되며, 추정된 특성에 따라 분류기(Linear Classifier)를 통해 어느 연령대에 속하는 지 분류되게 된다.
- [0072] 이를 위해, 학습 모델(220)은 다양한 연령대에 대한 오디오 데이터를 변환한 이미지 데이터를 이용하여 학습이 수행될 수 있으며, 여기서, 학습을 위한 오디오 데이터로는 공지의 라이브러리, 일례로서 'torchaudio' 라이브러리가 제공하는 데이터가 이용될 수 있다.
- [0073] 텍스트 변환모듈(230)은 실질적인 커뮤니케이션 서비스를 구현하기 위한 구성부로서, 대화 개시에 따른 발화자의 사용자 오디오 데이터(wav)를 입력받아 텍스트 데이터(text)로 변환할 수 있다.
- [0074] 이를 위해, 텍스트 변환모듈(230)은 오디오 데이터(wav)를 텍스트 데이터(text)로 변환하는 상용화된 API를 이용할 수 있으며, 일례로서, Google STT API(30)가 이용될 수 있다.
- [0075] 응답 모듈(250)은 입력되는 텍스트 데이터(text)에 대응하는 응답 데이터를 생성할 수 있고, 필터링 모듈(260)은 텍스트 데이터(text)에 포함되는 비속어를 필터링하거나 연령대에 부적합한 단어 또는 문장을 제거할 수 있다.
- [0076] 이를 위해, 응답 모듈(250) 및 필터링 모듈(260)은 비속어 탐지 모델을 포함하는 대화형 인공지능 플랫폼(50)을 이용할 수 있다. 이러한 대화형 인공지능 플랫폼(50)은 Text Tokenizing, 비속어 리스트 정의 및, 연령별 일상 대화 데이터 셋을 통한 비속어 탐지 모델의 학습 절차를 걸쳐 비속어 탐지 결과를 제공할 수 있고, 순화된 단어 정의, 비속어 순화 단어 매핑 학습 절차를 거쳐 텍스트 변환 결과를 제공할 수 있다.
- [0077] 도 5는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에서 대화형 인공지능 플랫폼을 통한 대화 생성 형태를 예시한 도면이다. 도 5를 함께 참조하면, 대화형 인공지능 플랫폼으로서 KoGPT2를 이용한 예로서, 입력되는 텍스트 데이터의 토큰과, 추정된 연령 레이블에 따라 출력 텍스트 데이터를 도출하는 형태를 예시하고 있다.
- [0078] 대화 출력 모듈(270)은 응답 모듈(250) 및 필터링 모듈(260)로부터 제공되는 답변 텍스트(text) 또는 필터링된 텍스트(text)를 상용화된 API를 이용하여 음성 데이터(wav)로 변환하여 대화 음성 데이터로서 출력할 수 있다. 이를 위한 API로는, Google TTS API(40)가 이용될 수 있다.
- [0079] 이를 통해, 대화 출력 모듈(270)은 제공되는 텍스트 데이터(text)를 TTS API에 입력하여 음성으로 변환하고, 변환된 음성을 가상객체의 답변 또는 아바타간 대화로서 출력하게 된다.
- [0080] 또한, 본 발명의 일 실시예에 따르면, 아바타 및 아바타간 또는, 아바타 및 가상객체간 대화에서, 음성에 대응하여 아바타 및 가상객체가 적합한 행동을 취하는 제스처 행동 기능을 더 제공할 수 있다.
- [0081] 이를 위한 구성부로서, 제스처 생성부(300)는 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하는 단계 및, 아바타 또는 가상객체의 대화에 따른 음성 출력시 제스처 데이터를 동기화하여 동시에 메타버스 구현부(도 2의 120)에 출력할 수 있다.
- [0082] 도 6은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 적용된 제스처 생성 기법을 예시한 도면이다.
- [0083] 도 6을 함께 참조하면, 본 발명의 일 실시예에 따르면 메타버스 공간내에서 사용자가 아바타 및 가상객체와의 자연스럽게 원활한 상호 작용을 가능하게 하기 위해, 제스처 생성부(300)는 Emotion Classification Model을 포함하는 'Gesticulator'를 사용할 수 있으며, 이는 각 음성의 음향 및 의미론적 표현을 입력으로 취하고 아바타

의 관절 각도 회전 등의 제스처를 구현할 수 있다.

- [0084] 이를 통해, 제스처 생성부(300)에 의해 출력된 제스처는 실시간으로 아바타 또는 가상객체에 적용될 수 있고, 이에 따라 보다 자연스러운 아바타간 대화가 가능해지게 된다.
- [0085] 사람은 말을 하면서 자연스럽게 대화의 맥락, 감정, 혹은 단어에 맞는 손짓이나 몸짓을 하게 되는데, 이는 소통과 정보 전달에 중요한 역할을 하는 행위라 할 수 있다. 특히, 실제 사람과 대면하는 것이 아닌, 메타버스 공간에서 자신을 대체하는 아바타로 커뮤니케이션을 수행하는 순간에는 이런 제스처의 역할이 더 중요하게 작용한다고 할 수 있다.
- [0086] 이하, 도면을 참조하여 전술한 구조의 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 커뮤니케이션 서비스를 제공방법을 상세히 설명한다.
- [0087] 도 7은 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 커뮤니케이션 서비스 제공방법을 나타낸 도면이다. 이하의 설명에서 각 단계별 실행주체는 별도의 기재가 없는 한, 본 발명의 시스템 및 그 구성부가 된다.
- [0088] 먼저, 본 발명의 시스템을 이용하여 메타버스 기반 커뮤니케이션 서비스를 이용하고자 하는 사용자는 사용자 단말을 통해 시스템에 접속하여 메타버스 공간에서 자신을 대체할 아바타를 생성할 수 있고, 시스템은 아바타를 메타버스 공간에 배치하여 사용자 단말에 제공하게 된다(S100).
- [0089] 여기서, 메타버스 공간은 시스템에 접속한 복수의 사용자 및 시스템에서 생성한 복수의 가상객체를 포함하고, 아바타들과 가상객체들이 대화를 비롯하여 쇼핑, 게임 등 여러 형태의 액티비티를 수행할 수 있는 공간을 가리킨다.
- [0090] 다음으로, 사용자 단말을 통해 아바타, 즉 사용자의 음성이 입력되면, 시스템을 그 음성을 이용하여 사용자의 연령을 추정할 수 있다(S110). 상세하게는, S110 단계는, 오디오 데이터를 입력받아 이미지 데이터로 변환하는 단계 및, 학습 모델의 CNN을 통해 이미지 데이터의 특징맵을 추출하고, 학습 모델의 분류기를 통해 특징맵에 기초하여 이미지 데이터에 해당하는 발화자의 연령을 복수의 그룹으로 분류하는 단계를 포함할 수 있다.
- [0091] 이를 위해, 학습 모델은 오디오 데이터를 이용한 훈련이 선행되어야 하며, 이러한 훈련 절차는, 변환기를 통해 음성 대화가 가능한 최소 연령에서부터 최대 연령을 대상으로 하는 다수의 발화자의 학습 오디오 데이터를 입력받는 단계 및, 오디오 데이터를 이미지 데이터로 변환하여 상기 학습 모델을 훈련하는 단계로 이루어질 수 있다. 이때, 상기 이미지 데이터로는 오디오 데이터를 시간, 주파수 및 세기의 3 요소로 표현한 Mel-Spectrogram이 이용될 수 있다.
- [0092] 다음으로, 커뮤니케이션 서비스로서 아바타 및 가상객체간 대화에서는 분류된 사용자의 연령에 따라 대화생성 모델을 이용하여 아바타의 발화 내용을 필터링하여 출력할 수 있고, 가상객체에 의한 연령에 적합한 응답을 생성할 수 있다(S120). 이를 위한 구체적 절차로서, 연령 레이블을 통해 아바타의 연령을 판단하는 단계, 텍스트를 대화형 인공지능 플랫폼에 입력하여 해당 연령대에 대응하는 답변 텍스트를 생성하는 단계 및, 답변 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 가상객체의 답변으로서 생성하는 단계를 포함할 수 있다.
- [0093] 또한, S120 단계 중, 아바타 및 아바타간의 대화에서는 분류된 사용자의 연령에 따라 대화생성 모델을 이용하여 두 아바타의 발화 내용을 필터링하여 출력할 수 있다. 이를 위한 구체적 절차로서, 음성을 변환한 텍스트에 연령 레이블을 매칭하는 단계, 연령 레이블을 통해 판단되는 연령별 일상대화 데이터셋을 이용하여, 비속어 탐지 모델을 통해 연령 레이블에 따른 사용자 연령별 텍스트에 포함된 비속어를 탐지하여 필터링하는 단계 및, 필터링된 텍스트를 TTS API에 입력하여 음성으로 변환하여 변환된 음성을 아바타간 대화로서 생성하는 단계를 포함할 수 있다.
- [0094] 다음으로, S120 단계를 통해 생성된 대화내용을 각 아바타 또는 가상객체의 음성으로 사용자 단말에 출력함으로써 실시간 대화를 구현하게 된다(S130).
- [0095] 한편, 본 발명의 일 실시예에 따르면, S130 단계에서 시스템은 대화 내용을 음성 뿐만 아니라, 아바타 및 가상객체의 제스처에 반영할 수 있으며, 이를 위한 단계로서, 사용자의 언어인 한국어 텍스트를 영어로 변환하는 단계, 변환된 텍스트를 Emotion Classification 모델에 입력하여 제스처 데이터를 생성하는 단계 및, 아바타 또는 가상객체의 대화에 따른 음성 출력시 제스처 데이터를 동기화하여 동시에 출력하는 단계를 포함할 수 있다.

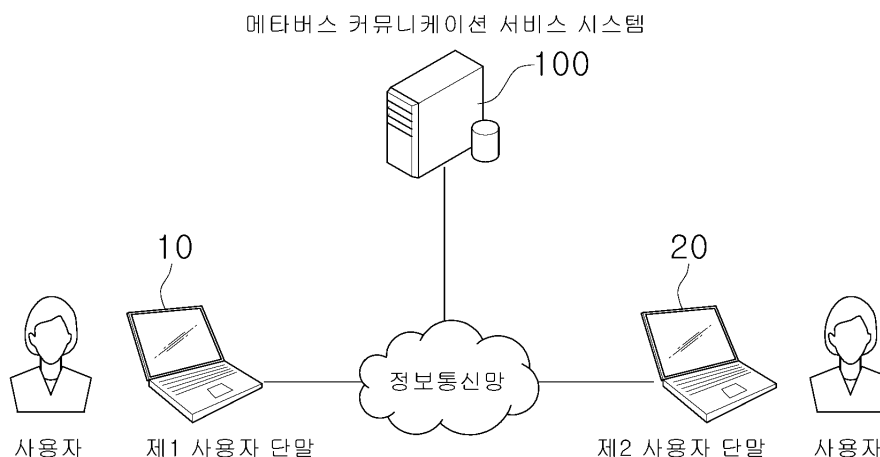
- [0096] 이하, 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 대한 테스트 결과를 통해 본 발명의 기술적 사상을 설명한다.
- [0097] 도 8 및 도 9는 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에 의한 연령 추정 정확도 및 대화 생성 모델에 의한 생성 결과를 예시한 도면이다.
- [0098] 도 8을 참조하면, 본 발명의 실시예에 따른 대화 생성 모델 기반 메타버스 커뮤니케이션 서비스 시스템에서 이용하는 음성을 통한 Age Classification model에 대하여, 'torchaudio' 라이브러리에서 제공하는 연령별 오디오 데이터를 통해 정확도를 측정한 결과 70% 이상을 확보함을 확인할 수 있다.
- [0099] 또한, 도 9를 참조하면, 대화 생성 모델에 의해 음성 데이터를 입력시, 예시된 바와 같이, 미리 설정된 비속어를 감지 및 필터링할 수 있고, 화자의 연령에 대응하여 부적절한 내용은 필터링을 처리가 가능하게 된다.
- [0100] 아울러, 오디오 데이터는 'child', 'adult' 및 'senior'의 연령대별 그룹으로 분류됨에 따라, 입력된 음성 데이터가 연령대에 적합한 형태의 문장으로 변환되어 출력됨을 확인할 수 있다.
- [0101] 상기한 설명에 많은 사항이 구체적으로 기재되어 있으나 이것은 발명의 범위를 한정하는 것이라기보다 바람직한 실시예의 예시로서 해석되어야 한다. 따라서, 발명은 설명된 실시예에 의하여 정할 것이 아니고 특허청구범위와 특허청구범위에 균등한 것에 의하여 정하여져야 한다.

부호의 설명

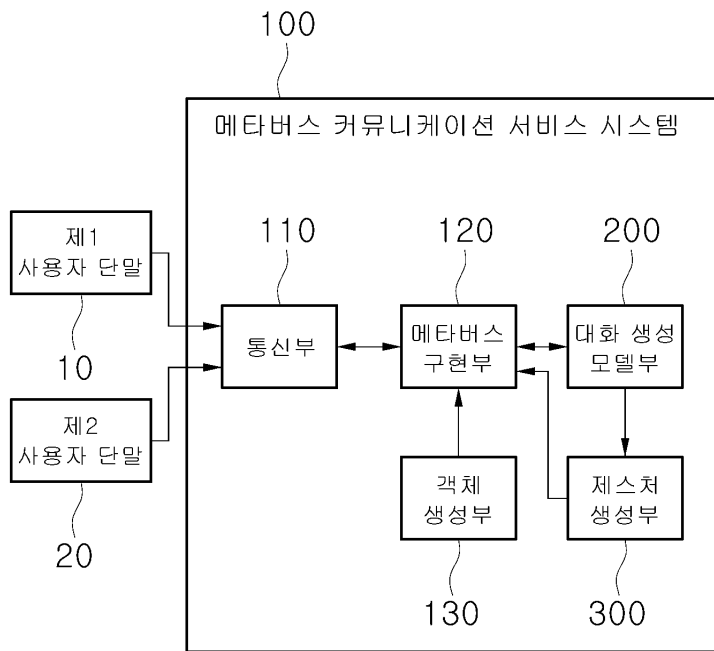
- [0102] 10, 20 : 사용자 단말 30 : STT API
40 : TTS API 50 : 대화형 인공지능 플랫폼
100 : 메타버스 커뮤니케이션 서비스 시스템
110 : 통신부 120 : 메타버스 구현부
130 : 객체 생성부 200 : 대화 모델 생성부
210 : 도메인 변환모듈 220 : 학습 모델

도면

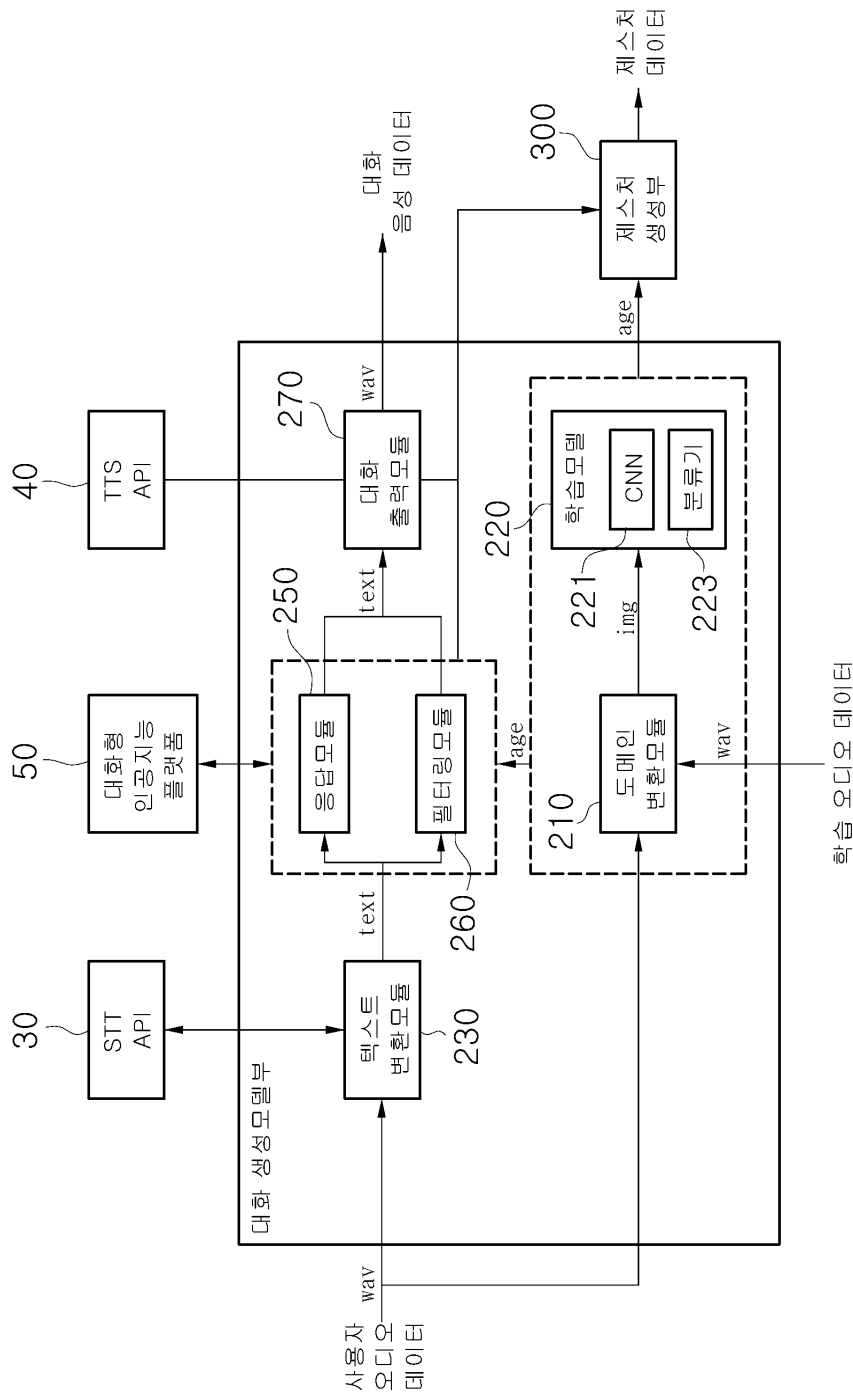
도면1



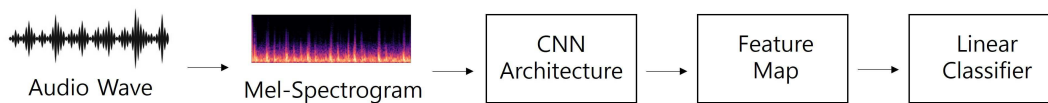
도면2



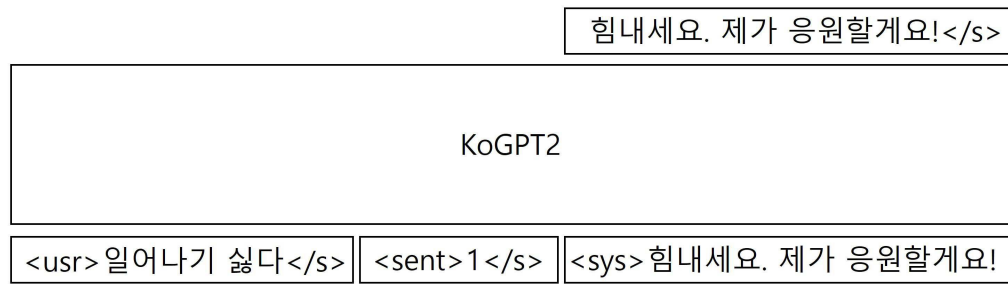
도면3



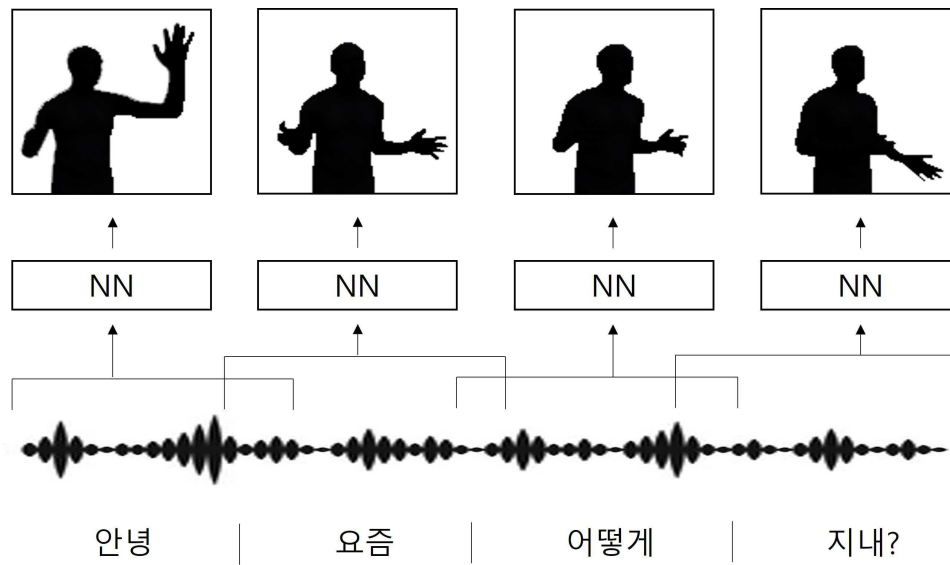
도면4



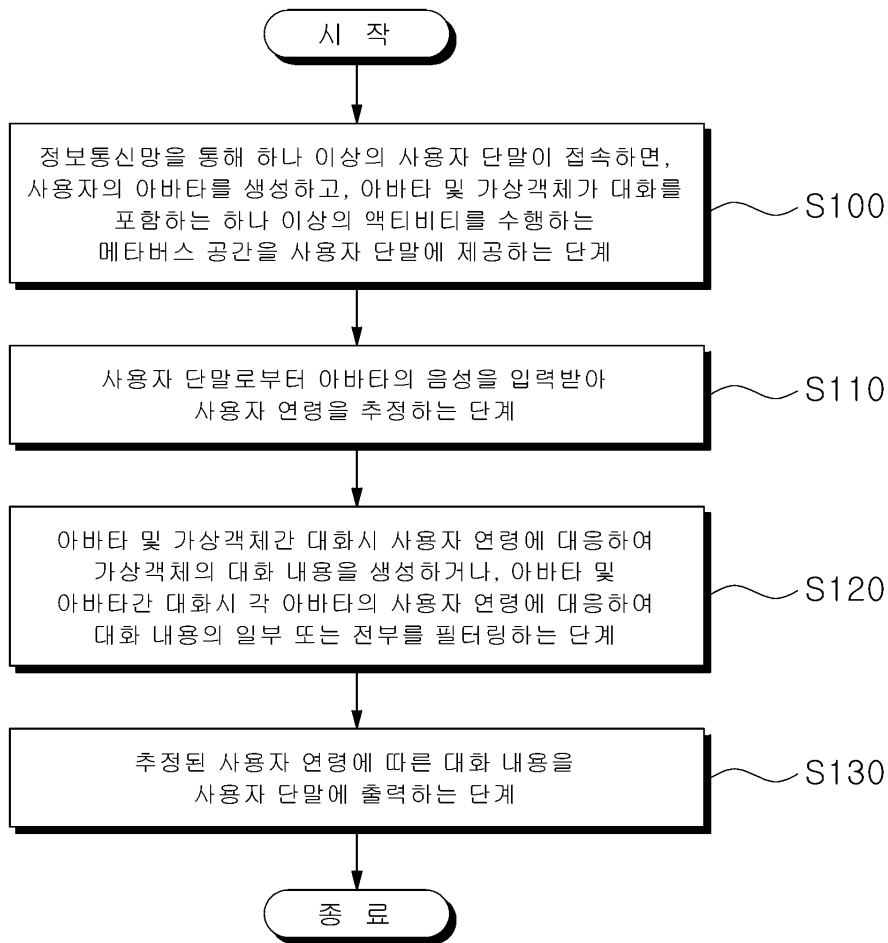
도면5



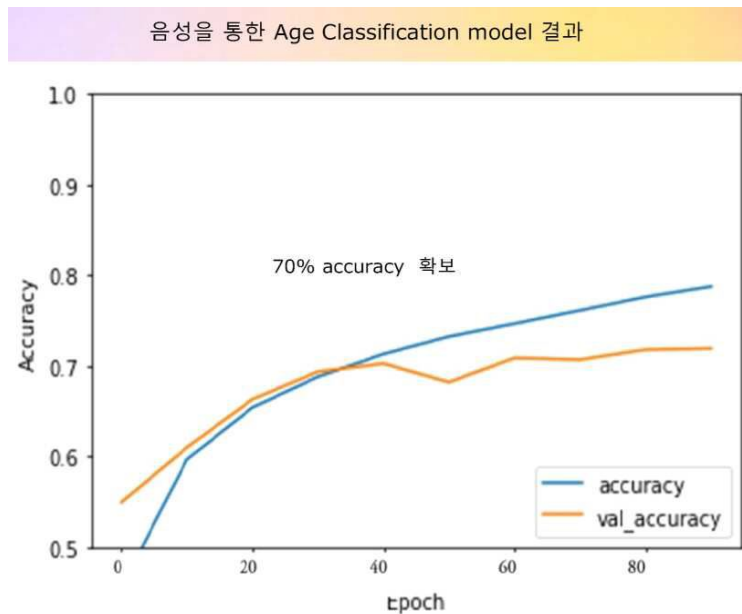
도면6



도면7



도면8



도면9

