Journal of KIIT. Vol. 24, No. 3, pp. 199-207, Mar. 31, 2026. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2026.24.3.199

VR 환경에서 LLM을 활용한 음성 제어 시스템의 설계와 구현

강인주*, 최현빈**[1], 김중락**[2], 유선진***

Design and Implementation of an LLM-based Voice Control System for Virtual Reality Environments

Inju Kang*, Hyunbin Choi**[1], Joongrock Kim**[2], and Sunjin Yu***

본 논문은 2025년도 교육부 및 경상남도의 재원으로 경상남도RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다.(2025-RISE-16-002) 또한, 2025년도 교육부의 재원으로 글로컬대학사업의 지원을 받아 수행된 사업의 결과입니다.

요 약

본 연구는 대규모 언어 모델(LLM) 기반 음성 인터페이스를 이용해 VR 환경에서 음성 명령으로 객체를 생성·조작하는 시스템을 제안하고, 사용자 실험으로 사용성을 평가하였다. Whisper-1 기반 STT와 GPT-4.1 기반 의미 해석을 결합해 사용자의 발화를 객체 제어용 JSON 명령으로 변환하였다. 생성된 JSON 명령은 Unity 기반 VR 환경에 적용되어 객체 생성 및 조작을 수행한다. Meta Quest 3를 활용한 VR 인테리어 시나리오에서 시스템 성능과 사용자 경험을 측정하였다. 실험 결과 평균 명령 처리 시간은 3.8초로 나타났다. 명령 성공률은 82.0%를 기록하였다. 시스템 사용성은 SUS 평균 75점으로 평가되었다. 이를 통해 LLM 기반 음성 인터페이스가 VR 객체 제어에 유효하게 적용될 가능성을 확인하였다.

Abstract

This study proposes an LLM-based voice interface that enables users to create and manipulate objects in a VR environment via speech commands, and evaluates its usability through a user study. We combine Whisper-1 for speech-to-text with GPT-4.1 for semantic interpretation to convert spoken utterances into JSON-formatted control commands. The generated JSON commands are executed in a Unity-based VR environment to support object creation and manipulation. A VR interior-design scenario was implemented and tested on Meta Quest 3 to measure system performance and user experience. Results show an average command processing time of 3.8 seconds. The command success rate reached 82.0%. Overall usability achieved a mean SUS score of 75. These findings demonstrate the feasibility of applying LLM-driven voice interfaces to VR object control.

Keywords

virtual reality, large language model, speech-to-text, object control commands

* 국립창원대학교 문화융합기술협동과정
- ORCID: https://orcid.org/0009-0001-8134-394X
** 국립창원대학교 인공지능융합공학과(**[2] 공동교신저자)
- ORCID[1]: https://orcid.org/0000-0003-4300-5859
- ORCID[2]: https://orcid.org/0009-0001-5284-1149
*** 국립창원대학교 메타융합콘텐츠학부, 인공지능융합공학과 교수(공동교신저자)
- ORCID: https://orcid.org/0000-0001-9292-4099

· Received: Dec. 29, 2025, Revised: Feb. 02, 2026, Accepted: Feb. 05, 2026
· Corresponding Author 1: Joongrock Kim
Dept. of Artificial Intelligence Convergence Engineering, 20 Changwondaehak-ro, Uichang-gu, Changwon-si, Gyeongsangnam-do, Korea
Tel.: +82-55-213-3962, Email: jrkim@changwon.ac.kr
· Corresponding Author 2: Sunjin Yu
Dept. of Artificial Intelligence Convergence Engineering, Meta-Convergence Content, Meta-Convergence Content Major, 20 Changwondaehak-ro, Uichang-gu, Changwon-si, Gyeongsangnam-do, Korea
Tel.: +82-55-213-3098, Email: sjyu@changwon.ac.kr

Ⅰ. 서 론

가상현실(VR, Virtual Reality)은 교육, 훈련, 설계, 엔터테인먼트 등 다양한 분야로 활용이 확대되고 있으며, 사용자와 가상 환경 간의 상호작용 방식은 VR 경험의 몰입도와 사용성을 좌우하는 핵심 요소이다[1]. 현재 VR 시스템에서는 컨트롤러 기반 인터페이스를 중심으로 사용되지만, 장시간 사용 시 신체적 피로를 유발한다는 한계가 있다[2].

이러한 문제를 보완하기 위한 대안으로 음성 기반 인터페이스가 주목받고 있다. 음성 입력은 직관적인 상호작용을 가능하게 하지만, 기존 시스템은 사전에 정의된 명령 위주로 동작하여 복합적인 문장이나 문맥을 이해하는 데 한계가 있다[3]. 최근 대규모 언어 모델(LLM, Large Language Model)의 발전은 이러한 한계를 극복할 수 있는 가능성을 제시한다[4]. LLM은 사용자의 발화를 의미 단위로 해석하고 의도를 추론할 수 있어, VR 환경에서 보다 자연스럽고 유연한 객체 제어를 가능하게 한다.

본 논문은 LLM 기반 음성 인터페이스를 활용한 VR 객체 제어 시스템을 설계·구현하는 것을 목적으로 한다. 제안하는 시스템은 음성을 텍스트로 변환한 후 LLM을 통해 행동, 객체, 공간 정보를 포함한 구조화된 명령으로 변환하고, 이를 VR 엔진에 전달하여 가상 환경에 반영한다. 본 연구의 기여는 LLM 기반 의미 해석 방식 제안, 음성 입력부터 VR 객체 제어까지의 전체 파이프라인 구현, 그리고 가상 인테리어 시나리오 기반 실험을 통한 실용성 검증에 있다.

Ⅱ. 관련 연구

2.1 VR 환경에서의 음성 인터페이스 연구

VR 환경의 사용자 인터페이스는 몰입감과 사용성을 좌우하는 핵심 요소로, 초기에는 핸드 컨트롤러와 같은 물리적 입력 장치를 중심으로 발전하였다. 컨트롤러 기반 인터페이스는 높은 조작 정밀도를 제공하지만, 반복적인 조작으로 인한 피로와 학습 부담이라는 한계를 지닌다[5]. 이러한 한계를 보완하기 위해 VR 환경에 음성 인식 기반 입력 방식을 적용한 연구들이 제안되었으며, 물리적 조작 없이도 상호작용이 가능함을 보여주었다[6]. 그러나 초기 연구들은 대부분 사전에 정의된 키워드 기반 명령에 의존하여, 자연스러운 발화나 복합적인 명령을 처리하는 데에는 제약이 있었다[7][8].

2.2 GPT 기반 자연어 인터랙션 연구

최근 LLM 모델을 활용한 자연어 인터랙션 연구가 VR 및 XR 환경을 중심으로 활발히 이루어지고 있다. 선행 연구들은 LLM이 자연어 명령으로부터 사용자 의도를 추론하고, 이를 가상 객체 생성이나 조작 명령으로 변환할 수 있음을 보여준다[9][10].

X. E. Wang et al.[11]의 연구에서는 LLM과 제스처 입력을 결합하여 사용자의 발화 및 가리킴(Pointing) 정보를 통합 해석함으로써, 자연스러운 복합 명령과 다중 객체 조작을 지원하였다. 사용자 실험에서 이 시스템은 기존 VR 조작 방식에 비해 피로도와 작업 부담을 감소시키는 것으로 나타났다.

또한 S. Parket al.[12]는 음성, 컨트롤러, 시각 정보를 LLM이 통합 분석하여 협업형 객체 조작과 로봇 제어를 지원하는 시스템을 제안하였다. 이러한 멀티모달 접근은 정밀한 작업 환경에서 음성 명령을 보조 입력으로 활용하는 효율성을 입증하였다.

Y. Xing et al.[13]은 자연어 발화와 제스처를 통해 아이디어를 시각화하도록 설계된 VR 시스템을 제안하여, 음성 입력이 창의적 작업의 몰입도를 높일 수 있음을 확인하였다.

2.3 관련 연구 분석 및 연구 공백

기존 연구를 종합하면, VR 환경의 음성 인터페이스는 주로 키워드 기반 명령이나 보조적 입력 방식에 머물러 왔으며, LLM 기반 접근 역시 대화형 에이전트나 멀티모달 시나리오에 초점을 두어 왔다. 이에 따라 음성 입력을 객체 조작의 주요 입력 수단으로 설정하고 자연 발화를 처리하는 인터페이스에 대한 정량적 검증은 충분히 이루어지지 않았다.

본 연구는 Whisper 기반 음성 인식과 GPT-4.1 기

반 자연어 이해를 결합한 VR 음성 인터페이스를 설계·구현하고, 명령 처리 시간과 성공률, 복합 발화 오류 특성에 대한 정량적 평가와 사용성 분석을 통해 그 적용 가능성과 한계를 검증하고자 한다.

III. 시스템 설계 및 구현

3.1 시스템 설계 개요

제안된 시스템은 Unity 2022.3.33f1(LTS) 기반의 VR 환경에서 동작하며, VR HMD는 Meta Quest 3를 사용한다. 자세한 개발 환경은 표 1에서 확인할 수 있다.

표 1. 개발 환경 및 시스템 구성 요소
Table 1. System development environment

Category	Description
Programming languages	Python 3.12.12, C#
Development platform	Unity 2022.3.33f1 (LTS), Visual Studio 2022, Jupyter Notebook 7.2.2
Libraries & frameworks	FastAPI, Uvicorn, Requests, JSON, Socket, Asyncio-related modules
API	GPT-4.1 API, Whisper-1 .API (Speech-to-Text)
VR device	Meta Quest 3

본 연구에서는 LLM 기반 음성 인터페이스를 활용하여 사용자의 자연어 음성 명령을 VR 환경 내 객체 제어로 연결하는 시스템을 설계하고 구현하였다. 그림 1은 제안하는 시스템의 전체 구성과 모듈 간 데이터 흐름을 나타낸다. 시스템은 음성 입력 및 인식 모듈, LLM 기반 명령 해석 모듈, 서버 연동 모듈, 그리고 VR 실행 모듈로 구성된다.

사용자의 음성 입력은 Meta Quest 3에 내장된 마이크를 통해 수집 후 음성 인식 모듈(Speech-to-Text, STT)을 통해 텍스트로 변환된다. 변환된 텍스트는 LLM 기반 명령 해석 모듈로 전달되어 의미 단위로 분석되며, 분석 결과는 구조화된 JSON 형식으로 Fast API 서버를 거쳐 VR 실행 모듈에 전달된다. VR 실행 모듈은 전달받은 명령을 Unity 기반 VR 환경에서 객체 생성 및 조작으로 반영한다.

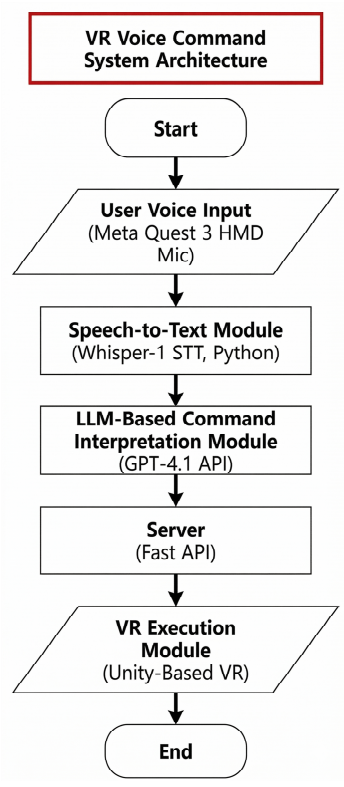

그림 1. 시스템 전체 구성도
Fig. 1. Overall system architecture

3.2 음성 명령의 의미 해석 및 구조화 구현

본 절에서는 사용자의 음성 명령을 해석하여 VR 실행 모듈에서 직접 활용 가능한 객체 제어 명령의 구조를 정의한다. 기존 VR 음성 인터페이스는 키워드 기반 또는 고정된 명령 패턴에 의존하여 복합적인 자연어 표현을 처리하는 데 한계가 있다. 본 연구에서는 이러한 한계를 보완하기 위해 대규모 언어 모델을 활용한 의미 기반 명령 해석 방식을 적용하였다.

음성 명령의 의미 해석을 위해 OpenAI의 GPT-4.1 모델을 사용하였다. 해당 모델은 문맥과 문장 구조를 고려한 의미 추론에 강점을 지니며, 다양한 자연어 표현을 안정적으로 해석할 수 있다. 또한 프롬프트 수준에서 출력 형식과 해석 규칙을 명확히 정의하여, 모델이 VR 객체 제어에 직접 사용 가능한 일관된 명령을 생성하도록 설계하였다. 프롬프트 규칙의 세부 내용은 표 2에 제시한다.

표 2. LLM 명령 해석을 위한 프롬프트 규칙
Table 2. Prompt rules for LLM command interpretation

Category	Rule description
Output format	JSON-only executable commands
Coordinate basis	Main camera coordinate system
Position definition	Explicit X, Y, Z values
Rotation	Axis-angle representation
Scale	Absolute scale values

본 시스템에서 처리 가능한 객체 제어 명령은 생성(Create), 이동(Move), 회전(Rotate), 크기 변경(Scale), 색상 변경(Change color)으로 제한된다. 각 명령은 단일 객체 또는 복수 객체에 대해 독립적으로 적용될 수 있으며, 이동 명령의 경우 거리 기반 상대 이동만을 지원한다. 반면, 물리 시뮬레이션 제어, 객체 간 충돌 조건 설정, 복잡한 연속 애니메이션과 같은 고급 상호작용 명령은 본 시스템의 범위를 벗어나므로 처리 대상에서 제외된다. 출력 형식 제약을 통해 LLM은 자연어 설명을 포함하지 않고 JSON 형식의 객체 제어 명령만을 생성한다.

예를 들어 “옷장을 생성하고 뒤로 0.5 m 이동”이라는 음성 명령은 객체 생성과 이동을 포함하는 복수의 제어 명령으로 해석되며, 그림 2와 같이 구조화된 JSON 데이터로 변환된다.

```
[Original response from GPT]
 [
  {
    "action": "create",
    "target": "closet",
    "objectData": {
      "prefab": "closet",
      "position": [
        -0.9689797,
        1.78389931,
        -0.147262752
      ]
    }
  },
  {
    "action": "move",
    "target": "closet",
    "reference": "camera",
    "direction": "back",
    "distance": 0.5
  }
]
```

그림 2. LLM 기반 객체 제어 명령 예시
Fig. 2. Example of LLM-based object control commands

LLM이 시스템에서 지원하지 않는 명령을 생성하거나, 필수 파라미터가 누락된 명령을 출력하는 경우 해당 명령은 실행 단계에서 필터링된다. 이 경우 시스템은 해당 명령을 무시하고, 실행 가능한 명령만을 순차적으로 처리하도록 설계하였다. 또한 명령 해석 결과가 객체 제어 사양을 만족하지 않을 경우, 오류 상태를 기록하여 후속 분석에 활용한다. 이러한 예외 처리 과정은 잘못된 자연어 해석으로 인한 시스템 오류를 방지하고, VR 환경에서의 안정적인 실행을 보장하기 위한 장치이다.

상대적 방향 표현을 포함하는 음성 명령은 메인 카메라 좌표계를 기준으로 해석되며, 각 방향 정보는 명시적인 X, Y, Z 축 값으로 변환된다. 이를 통해 사용자 시점 중심의 공간 표현이 시스템 내부 좌표계와 일관되게 매핑될 수 있도록 하였다.

객체 변환과 관련된 명령 역시 실행 환경에 직접 적용 가능한 형태로 구조화하였다. 회전 정보는 axis-angle 형식으로 정의되며, 크기 변경은 객체의 스케일 값을 명시적으로 지정하는 방식으로 처리된다. 색상 변경 명령은 HEX 색상 값으로 제한하여 실행 단계에서 추가적인 해석 과정이 필요하지 않도록 구성하였다. 또한 하나의 음성 명령에 여러 객체가 포함된 경우에도 안정적인 처리를 위해 객체별 독립 명령을 생성하는 규칙을 적용하였다. 이러한 구조화 전략은 복합 음성 명령 상황에서도 명령 간 충돌을 방지하고 예측 가능한 출력 결과를 제공하는 데 기여한다. 본 절에서는 이와 같이 음성 명령의 의미를 객체 제어 명령의 사양으로 정의하였으며, 해당 명령이 실제 VR 환경에서 어떻게 실행되는지는 다음 절에서 설명한다.

3.3 VR 실행 모듈 연동 및 객체 제어 구현

본 절에서는 3.2절에서 정의된 객체 제어 명령이 VR 실행 모듈에서 실제 객체 동작으로 수행되는 과정을 설명한다. LLM 기반 명령 해석 모듈에서 생성된 JSON 형식의 객체 제어 명령은 FastAPI 기반 서버를 통해 VR 실행 모듈로 전달되며, Unity 기반 VR 환경에서 처리된다.

자연어 기반 음성 명령은 사용자의 시점을 기준

으로 한 상대적 공간 표현을 포함하는 경우가 많다. 이에 본 연구에서는 명령 해석 단계와 실행 단계 간의 좌표 불일치를 방지하기 위해, 객체 제어 연산을 메인 카메라(Main camera) 좌표계를 기준으로 일관되게 수행하도록 설계하였다. 그림 3은 메인 카메라 기준 좌표계를 나타낸다.

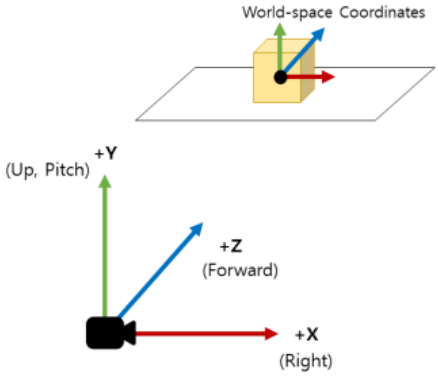

그림 3. 메인 카메라 기준 좌표계
Fig. 3. Main camera-based coordinate system

객체 이동 명령은 절대 위치 지정 방식이 아닌 상대 이동 방식으로 처리되며, 수신된 이동 값은 객체의 현재 상태에 증분(Delta) 형태로 누적 적용된다. 이러한 방식은 연속적이고 반복적인 음성 명령 입력 상황에서도 객체의 이동을 자연스럽게 이어지도록 한다.

회전 및 크기 조정 명령은 3.2절에서 정의된 axis-angle 및 scale factor 정보를 기반으로 Unity 엔진의 객체 변환 함수에 직접 매핑되어 실행된다. 이를 통해 추가적인 좌표 변환이나 재해석 과정 없이 LLM 출력 명령을 즉시 실행할 수 있다. 이와 같은 실행 구조를 통해 본 시스템은 자연어 기반 음성 명령을 사용자 시점 중심의 좌표 해석과 연계하여 VR 환경 내 객체 제어로 안정적으로 변환할 수 있다.

Ⅳ. 실험 및 결과

4.1 실험 설계 및 절차

본 연구에서는 GPT-4.1 기반 음성 인터페이스를 활용한 VR 객체 제어 시스템의 사용성과 성능을 검증하기 위해 사용자 실험을 수행하였다. 실험은 VR 인테리어 콘텐츠 환경을 기반으로 구성되었으며, 참가자들은 음성 명령을 사용하여 가상 공간 내 객체를 생성하고 이동·조정하는 과제를 수행하였다. 실험 시나리오는 실제 VR 환경에서 발생할 수 있는 객체 배치 상황을 단순화하여 구성되었으며, 음성 기반 상호작용의 직관성과 반응성을 평가하는 데 초점을 두었다.

실험에 사용된 콘텐츠는 가상 실내 공간에 객체를 배치하는 형태로 구성되었다. 그림 4는 본 연구에서 사용된 VR 인테리어 실험 환경을 보여주며, 사용자는 VR HMD를 착용한 상태에서 제한된 현실 공간 내에서 음성 명령을 통해 가상 실내 공간의 가구 객체를 생성하고 이동·조정한다. 본 실험은 가로 2 m, 세로 2 m의 제한된 현실 공간을 기준으로 수행되었으며, 해당 범위 내에서 사용자는 이동하면서 음성 기반 객체 제어를 수행할 수 있도록 설계되었다.

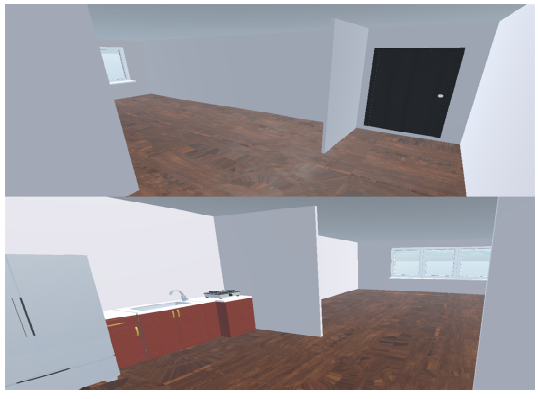

그림 4. VR 인테리어 실험 환경
Fig. 4. VR Interior design experiment environment

실험 참가자는 총 20명으로 구성되었고 이 중 15명은 VR 사용 경험이 없거나 제한적인 일반 성인이었고, 5명은 기존에 VR 사용 경험이 있는 참가자였다. 참가자들은 실험 시작 전 약 5분간의 사전 오리엔테이션을 통해 VR HMD 착용 방법과 기본적인 시스템 사용 방법에 대한 안내를 받았다. 이후 참가자들은 약 10분간 VR 공간에서 자유롭게 가구를 배치하며 음성 명령을 사용하여 객체 제어를 수행하였다. 실험 종료 후 참가자들은 시스템 평가를 위한 설문에 참여하였다.

4.2 평가 지표

본 연구의 실험 평가는 정량적 평가와 성적 평가를 병행하여 수행하였다. 정량적 평가는 명령 성공률과 명령 처리 시간을 기준으로 시스템의 성능을 측정한다. 정성적 평가는 사용성 설문 SUS(System Usability Scale)를 통해 사용자의 주관적인 사용 경험을 측정한다. SUS는 학습 용이성, 조작 편의성, 안정성 등의 측면을 평가하기 위한 총 10개 문항으로 구성되어있다. 표 3은 본 연구에서 사용한 SUS 설문 문항의 구성을 정리한 것이다[14].

표 3. SUS 설문 문항 구성
Table 3. SUS questionnaire items

Item	content
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person.
5	I found the various functions well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going.

각 문항은 5점 리커트 척도(1점은 '전혀 그렇지 않다', 5점은 '매우 그렇다')를 기준으로 응답을 수집한다. 실험에 사용한 설문은 한국어로 번역된 내용을 사용하였다[15].

4.3 실험 결과 및 분석

본 절에서는 제안한 GPT-4.1 기반 음성 인터페이스 시스템에 대한 사용자 실험 결과를 정량적 성능 평가와 사용성 평가를 중심으로 분석한다. 모든 결과는 참가자 20명의 실험 데이터를 기반으로 평균값과 표준편차를 산출하여 정리하였다. 먼저 시스템의 객관적인 성능을 평가하기 위해 음성 명령 처리 시간과 명령 성공률을 기준으로 정량적 분석을 수행하였으며, 해당 결과는 표 5에서 확인할 수 있다.

표 4. 정량적 성능 평가 결과
Table 4. Quantitative performance results

Metric	Mean	standard deviation	Unit
Command processing time	3,800	±700	ms
Command success rate	82.0	±7.0	%

음성 명령 입력 이후 객체 제어 결과가 VR 환경에 반영되기까지의 평균 처리 시간은 약 3.8초로 측정되었으며, 이는 음성 인식, GPT-4.1 기반 명령 해석, 서버 통신, Unity 기반 객체 제어 과정을 모두 포함한 엔드투엔드 처리 시간이다. 전체 명령의 평균 성공률은 82.0%(±7.0%)로 나타났고, 단순 객체 생성이나 이동과 같은 기본 명령에서는 비교적 안정적인 수행이 확인되었다.

총 200회의 음성 명령을 유형별로 분석한 결과, 단일 객체에 대한 단일 동작 명령이 전체의 약 62%를 차지했으며, 해당 유형의 성공률은 89.7%로 나타났다. 반면 두 개 이상의 동작이나 객체 참조를 포함한 복합·연속 명령은 전체의 38%를 차지했고, 이 경우 성공률은 79.5%로 상대적으로 낮았다. 이러한 결과는 발화에 포함된 의도 및 공간 정보가 증가할수록 자연어 해석 부담이 커지는 특성과 관련된 것으로 해석된다.

참가자 특성에 따른 비교에서는 VR 사용 경험이 있는 그룹의 평균 명령 성공률이 86.0%로, 비경험자 그룹의 78.0%보다 높게 나타났다. 이와 같은 차이는 특히 복합 명령 조건에서 두드러졌으며, VR 환경에 대한 사전 숙련도가 음성 기반 객체 제어 성능에 영향을 미칠 가능성을 시사한다.

기존 LLM 기반 VR 음성 인터페이스 연구와의 정량적 비교를 위해, 표 5에는 2장에서 분석한 선행 연구들의 성능 지표를 정리하였다. 앞서 2장에서 비교한 기존 연구를 살펴보면, Wang은 자연 발화 기반 멀티모달 인터페이스에서 약 3 - 5초 수준의 처리 시간을 보고한 반면, Park, Li, Zhan은 주로 대화형 또는 창의적 상호작용에 초점을 맞추어 명령 처리 시간이나 성공률과 같은 정량적 성능 지표를 명시적으로 제시하지 않았다. 이에 비해 본 연구는 음성 입력을 주된 상호작용 수단으로 설정한 상태에

서 명령 처리 시간과 성공률을 함께 측정함으로써, 기존 연구 대비 음성 기반 객체 제어 성능을 정량적으로 비교·분석했다는 점에서 차별성을 가진다.

표 5. 기존 VR 음성 인터페이스 연구와의 정량 비교
Table 5. Comparison with previous VR voice interface studies

Study	Input	Scope	Time	Eval
Wang et al.	Voice, Gesture	Multi-object	3 - 5s	Task-level
Park et al.	Voice, Controller	Multi-object	N/A	Qualitative
Li et al.	Voice, Gaze	Multi-object	N/A	Qualitative
Zhang et al.	Voice, Gesture	Creative ideation	N/A	Qualitative
Proposed	Voice	Object control	3.8s	82.0%

다음으로 시스템의 사용성을 평가하기 위해 SUS 설문을 실시하였다. 표 6은 SUS 설문 문항별 응답 결과를 평균 점수와 표준편차로 정리한 것이다.

표 6. SUS 문항별 응답 결과
Table 6. SUS item-wise results

Item	Mean	standard deviation	Item	Mean	Standard deviation
1	4.0	0.6	6	2.7	0.6
2	2.6	0.7	7	4.0	0.5
3	4.1	0.5	8	2.6	0.7
4	2.5	0.6	9	3.8	0.6
5	3.9	0.6	10	2.4	0.6

SUS 설문 결과, 사용 빈도와 사용 용이성을 평가하는 문항 1과 3은 각각 평균 4.0(±0.6), 4.1(±0.5)로 나타났으며, 기능 통합성과 학습 용이성을 평가하는 문항 5와 7 역시 평균 3.9(±0.6), 4.0(±0.5)의 점수를 기록하여 전반적으로 긍정적인 평가가 이루어졌다. 반면, 시스템 복잡성과 사용 부담을 묻는 역문항들은 평균 2.4~2.7 범위로 나타나 과도한 복잡성은 인식되지 않았으나 개선 여지가 있음을 시사하였다. 사용 자신감을 평가하는 문항 9는 평균 3.8(±0.6)로, 반복 사용을 통한 숙련도 향상 가능성을 보여주었다. 문항별 점수를 환산한 전체 평균 SUS 점수는 약 75점으로 기준 점수인 68점을 상회하여, 본 시스템이 전반적으로 수용 가능한 수준의 사용성을 제공함을 확인하였다.

평균 명령 처리 시간 3.8초와 명령 성공률 82.0%는 기존 LLM 기반 VR 음성 인터페이스 연구와 유사한 수준으로 나타났으며, 자연 발화 기반 객체 제어가 VR 환경에서도 실용적으로 적용 가능함을 시사한다. 특히 키워드 기반 명령이 아닌 자유 발화를 사용했음에도 비교 가능한 성능을 유지하였다는 점에서 의미가 있다.

한편 명령 유형에 따라 성능 차이가 확인되었다. 단일 객체 단일 동작 명령에서는 안정적인 수행이 이루어진 반면, 복합·연속 명령이나 복수 객체 명령에서는 오류 발생 비율이 증가하였다. 이는 발화에 포함된 의도 수와 공간 정보가 증가할수록 자연어 해석 부담이 커지기 때문으로 해석되며, 전체 200회의 음성 명령 분석에서도 동일한 경향이 나타났다. 또한 VR 사용 경험이 있는 참가자는 비경험자에 비해 더 높은 명령 성공률과 짧은 처리 시간을 보여, 사용자 숙련도가 음성 기반 인터페이스 활용 효율에 영향을 미칠 수 있음을 확인하였다.

사용성 측면에서 SUS 평균 점수는 75점으로 기본적인 수용 가능 수준의 사용성을 확보하였다. 다만 제한된 실험 공간과 치수 시각화가 제공되지 않은 환경에서는 거리 단위 음성 명령만으로 정밀한 위치 조정에 한계가 있었다. 이러한 특성을 고려할 때, 향후 연구에서는 명령어 유형과 오류 발생 특성, 그리고 참가자 유형에 따른 성능 차이를 보다 정량적으로 분석할 필요가 있다.

Ⅴ. 결론 및 향후 과제

본 연구는 GPT-4.1 기반 음성 인터페이스를 VR 환경에 적용하여, 자연어 음성 명령을 통해 가상 객체를 제어할 수 있는 시스템을 설계·구현하였다. 음성 및 의미 기반 해석을 통해 기존 컨트롤러 중심 VR 인터페이스의 학습 부담과 물리적 피로 문제를 보완하고자 하였다.

사용자 실험 결과, 평균 명령 처리 시간은 3.8초, 명령 성공률은 82.0%로 나타났으며, SUS 평균 점수는 75점으로 수용 가능한 수준의 사용성을 확인하

였다. 이는 음성 기반 인터페이스가 VR 환경에서 객체 생성 및 대략적인 위치 조정에 적용 가능함을 보여준다. 다만 복합적이고 연속적인 음성 명령에서는 해석 지연이나 오류가 발생할 수 있으며, 치수 시각화가 제공되지 않은 실험 환경에서는 거리 단위 음성 명령만으로 정밀한 위치 조정에 한계가 있었다. 향후 연구에서는 치수 시각화 및 정밀도 향상을 위한 보조 피드백 도입과 다양한 VR 콘텐츠로의 확장을 통해 시스템의 일반화 가능성을 검증할 예정이다.

References

[1] R. A. Putawa and D. Sugianto, "Exploring user experience and immersion levels in virtual reality: A comprehensive analysis of factors and trends", International Journal of Research on Metaverse, Vol. 1, No. 1, pp. 20-39, Jun. 2024. https://doi.org/10.47738/ijrm.v1i1.3.

[2] X. Lou, Q. Zhao, Y. Shi, and P. Hansen, "Arm posture changes and influences on hand controller interaction evaluation in virtual reality", Applied Sciences, Vol. 12, No. 5, pp. 1-22, Dec. 2022. https://doi.org/10.3390/app12052585.

[3] M. E. Hajji, T. A. Baha, A. Berka, H. A. Nacer, and H. E. Aouifi, "An architecture for intelligent tutoring in virtual reality: Integrating LLMs and multimodal interaction for immersive learning", Information, Vol. 16, No. 7, pp. 1-20, Jun. 2025. https://doi.org/10.3390/info16070556.

[4] Y. Tang, J. Situ, A. Y. Cui, M. Wu, and Y. Huang, "LLM integration in extended reality: A comprehensive review of current trends, challenges, and future perspectives", Proc. ACM Symp. on Applied Computing, Yokohama Japan, pp. 2012-2024, Apr. 2025. https://doi.org/10.1145/3706598.3714224.

[5] L. Cordioli, L. Piro, M. Valoriani, and M. Matera, "Exploring LLM-driven interaction for knowledge retrieval in extended reality", Proc. ACM Symp. on Virtual Reality Software and Technology, Salerno, Italy, pp. 112-123, Oct. 2025. https://doi.org/10.1145/3750069.3750160.

[6] D. Wang and C. Lee, "Low-Cost Implementation and Entrepreneurial Application of a VR-Based Digital Human Interaction System: A Design-Oriented Approach Using Unity and Google APIs", Journal of Venture Startup Research, Vol. 20, No. 3, pp. 217-232, Jun. 2025. http://dx.doi.org/10.16972/apjbve.20.3.202506.217.

[7] N. S. Saravanan and H. Shankar, "VoxGuru: A multimodal AI for transformative conversational learning in real-time dynamic environments", Proc. 9th Int. Conf. on Inventive Systems and Control (ICISC), Coimbatore, India, pp. 1478-1485, Aug. 2025. https://doi.org/10.1109/ICISC65841.2025.11187982.

[8] E. Dritsas, M. Trigka, C. Troussas, and P. Mylonas, "Multimodal interaction, interfaces, and communication: A survey", Multimodal Technologies and Interaction, Vol. 9, No. 1, pp. 1-32, Jan. 2025. https://doi.org/10.3390/mti9010006.

[9] M. Z. Afzal, S. A. Ali, D. Stricker, P. Eisert, A. Hilsmann, D. Perez-Marcos, and M. Cuadros, "Next generation XR systems: Large language models meet augmented and virtual reality", IEEE Computer Graphics and Applications, Vol. 45, No. 1, pp. 43-55, Feb. 2025. https://doi.org/10.1109/MCG.2025.3548554.

[10] X. Hu, D. Ma, F. He, Z. Zhu, S. K. Hsia, C. Zhu, and K. Ramani, "GesPrompt: Leveraging co-speech gestures to augment LLM-based interaction in virtual reality", Proc. ACM Designing Interactive Systems Conf., Madeira, Portugal, pp. 59-80, Jul. 2025. https://doi.org/10.1145/3715336.3735769.

[11] X. E. Wang, Z. P. Sin, Y. Jia, D. Archer, W. H. Fong, Q. Li, and C. Li, "Can You Move These Over There? An LLM-Based VR Mover for Supporting Object Manipulation", arXiv preprint arXiv:2502.02201, pp. 1-64, Feb. 2025. https://doi.org/10.48550/arXiv.2502.02201.

[12] S. Park, C. C. Menassa, and V. R. Kamat,

"Integrating Large Language Models with Multimodal Virtual Reality Interfaces to Support Collaborative Human-Robot Construction Work", Journal of Computing in Civil Engineering, Vol. 39, No. 1, Art. No. 4024053, pp. 1-39, Jan. 2025. https://doi.org/10.48550/arXiv.2404.03498.

[13] Y. Xing, J. Ban, T. D. Hubbard, M. Villano, and D. Gomez-Zara, "Immersed in My Ideas: Using Virtual Reality and Multimodal Interactions to Visualize Users' Ideas and Thoughts", arXiv preprint arXiv:2409.15033, pp. 1-24, Sep. 2024. https://doi.org/10.48550/arXiv.2409.15033.

[14] J. Brooke, "SUS: a 'quick and dirty' usability", Usability evaluation in industry(Taylor and Francis), pp. 189-194, Jun. 1996.

[15] Y. J. Lee and K. T. Jung, "Analysis of User Experience for the Class Using Metaverse-Focus on 'Spatial'", Journal of Practical Engineering Education, Vol. 14, No. 2, pp. 367-376, Aug. 2022. https://doi.org/10.14702/JPEE.2022.367.

저자소개

강 인 주 (Inju Kang)

2024년 8월 : 국립창원대학교 문화테크노학과(학사)
2024년 9월 ~ 현재 : 국립창원대학교 문화융합기술협동과정 석사과정
관심분야 : 대형언어모델, 가상현실, 실감콘텐츠

최 현 빈(Hyunbin Choi)

2021년 8월 : 국립창원대학교 문화테크노학과(학사)
2025년 2월 : 국립창원대학교 문화융합기술 협동과정(석사)
2025년 3월 ~ 현재 : 국립창원대학교 인공지능융합공학 박사과정
관심분야 : 컴퓨터비전,증강/가상현실, 모션캡처

김 중 락 (Joongrock Kim)

2005년 3월 : 고려대학교 전자정보공학(공학사)
2005년 1월 ~ 2006년 7월 : 엠텍비전 연구원
2008년 8월 : 연세대학교 생체인식공학(공학석사)
2014년 2월 : 연세대학교 전기전자공학(공학박사)
2014년 2월 ~ 2024년 12월 : LG전자 CTO인공지능연구소 Vision Intelligence 연구실 팀장
2025년 1월 ~ 현재 : 국립창원대학교 인공지능융합공학과 부교수
관심분야 : 2D/3D 컴퓨터비전, 인공지능, On-Device

유 선 진 (Sunjin Yu)

2005년 3월 : 고려대학교 전자정보공학(공학사)
2006년 2월 : 연세대학교 생체인식공학(공학석사)
2011년 2월 : 연세대학교 전기전자공학(공학박사)
2011년 1월 ~ 2012년 5월 : LG전자기술원 미래IT융합연구소 선임연구원
2012년 5월 ~ 2013년 2월 : 연세대학교 전기전자공학과 연구교수
2013년 3월 ~ 2016년 8월 : 제주한라대학교 방송영상학과 조교수
2016년 9월 ~ 2019년 8월 : 동명대학교 디지털미디어공학부 부교수
2019년 9월 ~ 2024년 9월 : 국립창원대학교 문화테크노학과 부교수
2024년 10월 ~ 2025년 2월 : 국립창원대학교 문화테크노학과 정교수
2025년 3월 ~ 현재 : 국립창원대학교 메타융합콘텐츠학부/인공지능융합공학과 정교수
관심분야 : 컴퓨터비전, 증강/가상현실, HCI