Behaviour Modeling of Virtual Autonomous Driving Agent Using Voice Command in Risky Scenarios

Velichko Minev, Dilyana Budakova

Introduction

- Autonomous driverless electric vehicles, such as those by Waymo [1], Zoox's robotaxis [2], AutoX's self-driving grocery delivery vehicles [3], and Tesla's Full Self-Driving (Supervised) [4],
- combine exciting innovative technologies aimed at ensuring a safe, environmentally friendly, and accessible urban mobile environment for people.

Introduction

- However, risky scenarios exist—such as encountering traffic jams, dense fog, smoke, fire, destruction, or rescue operations—where the automated system cannot independently make a decision on how to act.

- In these cases, various approaches are supported where a human assistant can provide remote commands. The system may switch to manual or remote control with a tele-operator, or operate semi-autonomously. While a tele-assistant only provides instructions, a human tele-operator can execute the driving task remotely, either fully or partially.

Conducting repeated tests of autonomous driving systems in risky scenarios is dangerous, expensive, and virtually impossible.

- **Realistic test platforms based on virtual reality and 3D modeling are used for these purposes. Examples include:**
- **Carcraft software developed by Google's Waymo automated driving team;**
- **AirSim system for autopilot vehicle testing by Microsoft [6];**
- **The Apollo virtual driving platform created by Baidu Apollo [7].**

Abstract

This is why it is of great relevance to model and study the behavior of virtual autonomous vehicles in risky scenarios and to seek solutions for improving this behavior.

This article models the behavior of a multimodal virtual agent—a vehicle—under conditions of reduced visibility in an urban environment.

In situations where the virtual agent cannot decide how to act, it receives voice commands from a human assistant.

Abstract

The vehicle-agent reasons, interprets, and is capable of executing voice instructions.

It operates in a hybrid control mode, which includes autonomy with the possibility for human voice intervention.

Abstract

An experimental evaluation has been conducted to assess the effectiveness of voice command recognition and interpretation by a system based on a Large Language Model (LLM) in risky scenarios with reduced visibility.

The results show that the agent's behavior improves as a result of receiving and executing voice commands from a remote operator-assistant in a risky environment.

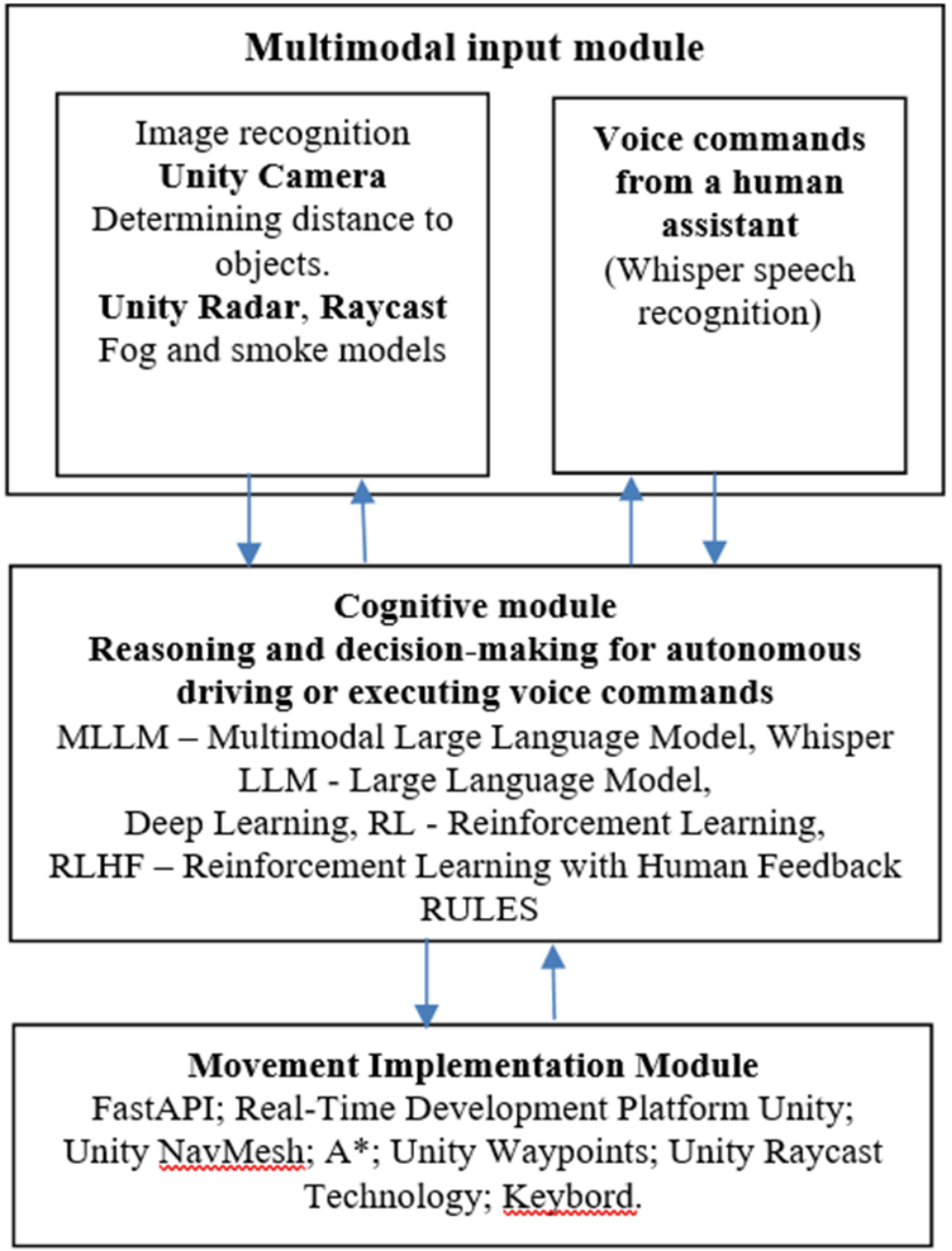


Figure 1. Architecture of a multimodal autonomous agent-driver.

The architecture of a multimodal virtual autonomous vehicle-agent integrates:

- Visual perception;
- Recognition of dynamic objects—people and other vehicles;
- Interpretation, reasoning, and execution of voice-given commands via Whisper and Large Language Model (LLM).

The model combines several current scientific fields:

- Multimodal autonomous agents;
- **Human-computer interaction via voice guidance**;
- 3D-modeled virtual environment as an experimental platform; ;
- Voice control under conditions of reduced visibility (fog or smoke);
- Voice control implemented using Whisper and a Large Language Model (LLM)

The workflow is organized into a sequential chain.

First, the voice command is captured and transcribed into text by Whisper.

This text, along with the visual data from the sensors in Unity, is fed into an MLLM, which generates a semantic understanding of the situation (e.g., recognizing that the obstacle is a truck).).

Subsequently, the LLM module, trained via RLHF, selects the safest action based on the provided command.

Finally, this decision is converted into specific vehicle control commands within the simulation..

Switching Between Autonomous Control and Voice Intervention

- **The tele-assistant continuously monitors the autonomous vehicle and assists it with voice commands.**

- **A safety threshold has been introduced, which is determined by the visibility in fog.**

- **In clear weather, the autonomous vehicle prefers to operate autonomously and maintains a high safety level.**

- **When visibility drops to six meters or less, the agent's safety level becomes low, and it begins to execute all provided voice commands.**

- **In a critical situation, without the help of the tele-assistant, the autonomous vehicle cannot continue its movement.**
- **At this stage of the system's implementation, the autonomous vehicle has full confidence in the tele-assistant.**

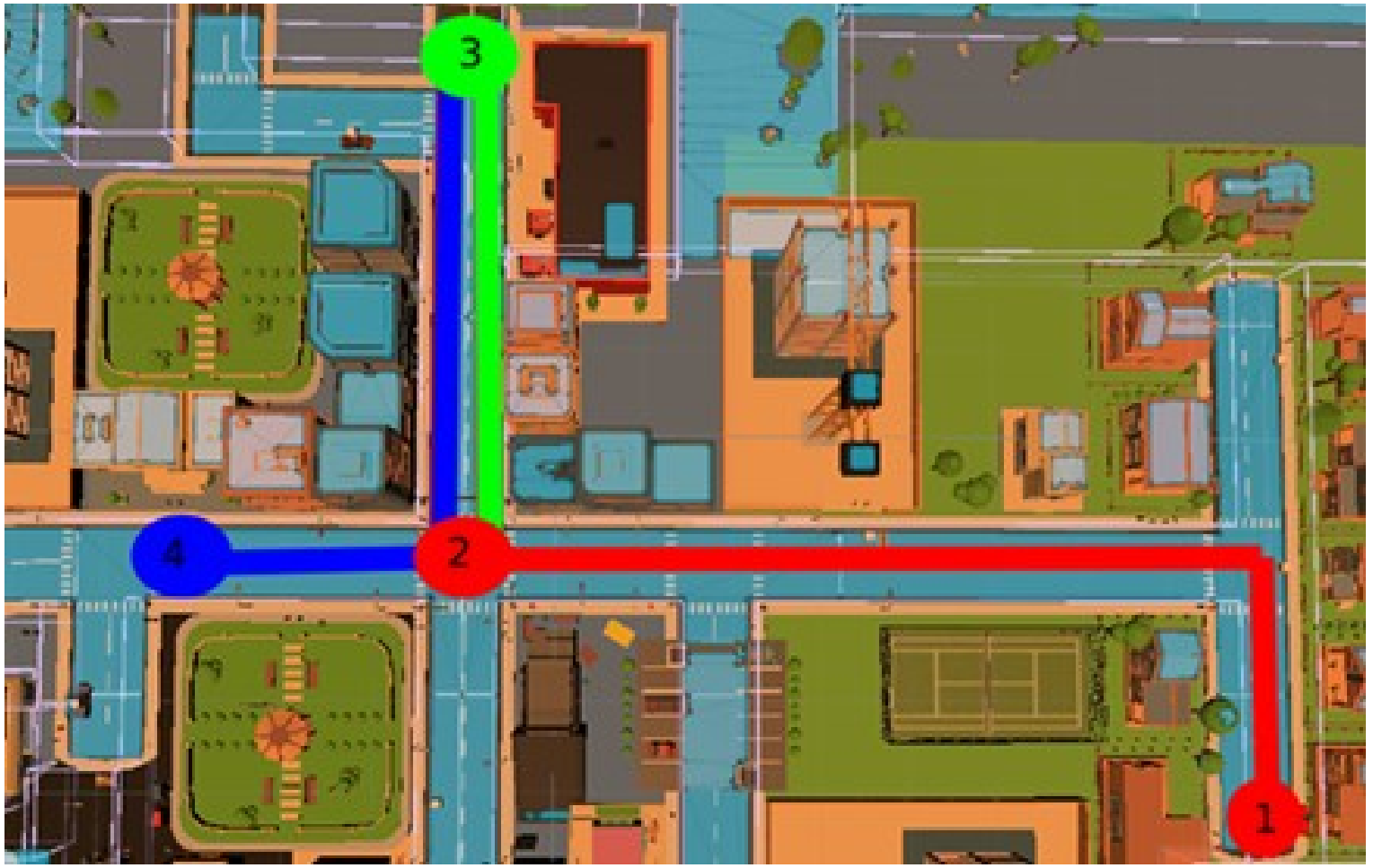
- **One possibility for future development of the system is to introduce an additional safety module into the cognitive architecture.**

- **This module could verify whether the voice command corresponds to the context received from the sensors.**

- **Depending on this, the command may or may not be executed.**

Urban environment model.

Experimental route marked with colored lines and numbering.



RESULTS OBTAINED FROM THE BEHAVIOR OF A VIRTUAL VEHICLE-AGENT UNDER CHANGING VISIBILITY..

Meteorological conditions. Degree of Visibility simulated with raycast	Time to travel on the given route	Need for voice commands to continue the agent's movement along the route
Sunny Rays Visibility – 20 m.	80,36 sec.	no
Foggy Rays Visibility – 8 m.	80,47 sec.	no
Foggy Rays Visibility – 6 m.	107,72 sec.	yes
Foggy Rays Visibility – 6 m.	109,23 sec.	yes
Foggy Rays Visibility – 4 m.	117,07 sec.	yes
Foggy Rays Visibility т – 4 m.	123,63 sec.	yes

The results indicate that under reduced visibility of approximately six (6) meters or less, for example four (4) meters, there is a necessity for driving assistance via voice commands.

RESULTS OF THE VIRTUAL VEHICLE-AGENT BEHAVIOR DURING DRIVING UNDER VARIOUS METEOROLOGICAL CONDITIONS. REACTION TIME FOR THE EXECUTION OF VOICE COMMANDS. NUMBER OF UNRECOGNIZED VOICE COMMANDS.

Meteorological conditions	Voice command response time	Route travel time	Number of unrecognized commands
Sunny	2,66 sec.	105,95 sec.	1 num.
Sunny	2,68 sec.	96,76 sec.	0 num.
Sunny	2,64 sec.	98,77 sec.	0 num.
Foggy	2,69 sec.	109,76 sec.	0 num.
Foggy	3,00 sec.	123,85 sec.	2 num.
Foggy	3,21 sec.	117,85 sec.	0 num.

The results show that there is a high recognition rate of voice commands.

Conclusion

- The results demonstrate that in risky scenarios, tele-assistance of autonomous driver systems with voice commands is highly effective and essential.

- Voice commands are utilized as high-level semantic control and reduce uncertainty instead of replacing low-level vehicle control.

Conclusion

- An advantage is the use of a large language model, such as the Unity Whisper model, to facilitate communication via voice commands.
- These LLMs allow for the use of free natural language expression during communication.
- The given commands can be interpreted correctly by the autonomous vehicle without requiring humans to learn a specific set of commands.

Conclusion

- Future iterations of the system will address potential LLM risks, such as hallucinations, by incorporating a cross-verification safety module to validate commands against real-time sensor data.

- Additionally, a robust fail-safe protocol is planned to ensure a controlled vehicle stop in the event of connectivity or model failure.

Thank you for your attention!