

Keep Gesturing: A Game for Pragmatic Communication

János Adrián GULYÁS^{a,1}, Miklós Máté BADÓ^{a,1}, Kristian FENECH^{a,2} and András LŐRINCZ^a

^aDepartment of Artificial Intelligence, Eötvös Loránd University, Hungary

Abstract. We present an innovative augmented reality game which aims to demonstrate a new dimension of interaction between humans and large language models through non-verbal gesture-based communication. Players collaborate with an LLM-controlled avatar to identify and correct discrepancies in an augmented reality environment, relying solely on non-verbal cues. The game aims to explore the emergence and evolution of pragmatic communication between humans and AI agents.

Keywords. Gesture-Based Communication, Human-AI Interaction, Large Language Model, HoloLens 2, Pragmatic Communication

1. Introduction

Human-machine interaction may occur across many types of communication channels [1], a common form for this channel to take is of natural language. However, natural language in the form of speech and text may not be suitable given restrictions set by the environment such as in noisy and dynamic environments where these modalities may lead to delays and possible misunderstanding. In such situations non-verbal communicative gestures present themselves as a robust alternative communication channel to enable rapid communication. Importantly the meaning behind individual gestures can change depending on the nature and context surrounding the interaction [2].

Our proposed game[‡] introduces a novel approach to human-machine interaction through gesture-based communication within a game developed for HoloLens 2. While recent work has explored the idea of collaboration between humans and large language models (LLMs), the focus has been on the text domain for example [3,4]. This game poses various collaborative challenges to a human player and a companion driven by a LLM, that requires them to communicate using exclusively hand gestures. The challenges aim to aid the development of a gesture-based pragmatic language between the two participants and will explore how this new language emerges and improves during various play-throughs with different players, and the implications of this phenomenon for the future of human-computer interaction.

¹Equal contributors

²Corresponding Author: Kristian Fenech, fenech@inf.elte.hu

[‡]Demo available https://youtu.be/j_bAw8e01NU?si=wVB23530QID6bMbU

2. Methodology

Although the game is specifically engineered for human-machine interaction, two humans are also able to complete the game in its entirety. The game takes place in augmented reality, with various interactive objects (puzzle elements) placed in different parts of the room. Interacting with these objects changes their properties and configuration. An object has a specific configuration that it must be set to at the end of the level for a point to be scored. However, only one of the two participants is able to interact with the objects in augmented reality, and only the other party has access to the rules that define the correct configurations. The two parties are not allowed to exchange this information verbally and are restricted to the use of gestures only.

The game features gesture-based back-and-forth interaction between human and a LLM. The hand-gesture approach allows for enough complexity so that a dictionary of gestures that represent various concepts might emerge. Various metrics are calculated, with some being presented to the participants as they complete (or fail) a level. This serves the purpose of encouraging the advancement of their emerging language, as better metrics reflect more effective communication. The presented metrics include efficiency and score, while hidden metrics also include the overall speed of completion, as well as the appearance of complex concepts in the language.

In our implementation, LLMs are instrumental by stepping into the role of one of the players, engaging in the game's two-way gesture-based communication. The model currently in use for this is OpenAI's GPT-4 [5]. After receiving an encoded series of hand motions from the players, the LLM interprets these inputs and replies in the same manner. This creates an environment where both parties can use each other's gestures and ideas to come to an understanding.

An animated character in augmented reality represents the LLM within the game's virtual environment, serving as its avatar. Gestures received from the LLM are visually depicted through this avatar's animation. By controlling its avatar, the LLM can intuitively reply to the human player, guiding them to identify and rectify the errors within the virtual room.

We aim to demonstrate that LLMs can understand and contribute to the creation of a new, personalized language, created in collaboration with a human. We believe this will pave the way for more unique and adaptive methods of interacting with artificial intelligence.

3. Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 952026. Additional support provided by the the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory.

References

- [1] Oviatt S, Cohen PR. The paradigm shift to multimodality in contemporary computer interfaces. Springer Nature; 2022 Jun 1.

- [2] Wharton T. *Pragmatics and non-verbal communication*. Cambridge University Press; 2009 Sep 17.
- [3] Rastogi C, Tullio Ribeiro M, King N, Nori H, Amershi S. Supporting human-ai collaboration in auditing llms with llms. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* 2023 Aug 8 (pp. 913-926).
- [4] Memmert L, Tavanapour N. Towards human-AI-collaboration in brainstorming: empirical insights into the perception of working with a generative AI.
- [5] OpenAI. (2023). ChatGPT (gpt-4-1106-preview) [Large language model]. <https://chat.openai.com/chat>