



Designing Non-Humanoid Virtual Assistants for Task-Oriented AR Environments

Bettina Schlager

Steven Feiner

Department of Computer Science, Columbia University
{b.schlager, feiner}@columbia.edu

Motivation

Embodied conversational agents in Augmented Reality have been shown to enhance social presence and reduce mental workload. However, **humanoid agents could introduce social biases and distractions**. For task-focused AR environment, we propose a non-humanoid virtual assistant design.

Method

We use a **simple geometric shape** for a non-humanoid agent, minimizing visual intrusion for user tasks, based on guidelines from previous literature [1,2]. Our prototype is implemented using Microsoft HoloLens 2.

Implementation

Characteristic

Non-intrusive Behavior (NB)

Natural Interactions (NI)

Just-In-Time Guidance (JIT)

Error Assistance (EA)

Directions (D)

Persistent Availability (PA)

Spatial Awareness (SA)

The agent, represented by two concentric rings, employs view management [3] for minimal task interference, and users can reposition it in 3D space if needed.

We use a predefined invocation phrase for user-agent communication, OpenAI GPT-3 for text responses, and Mixed Reality Toolkit 2.7 for spatialized audio replies.

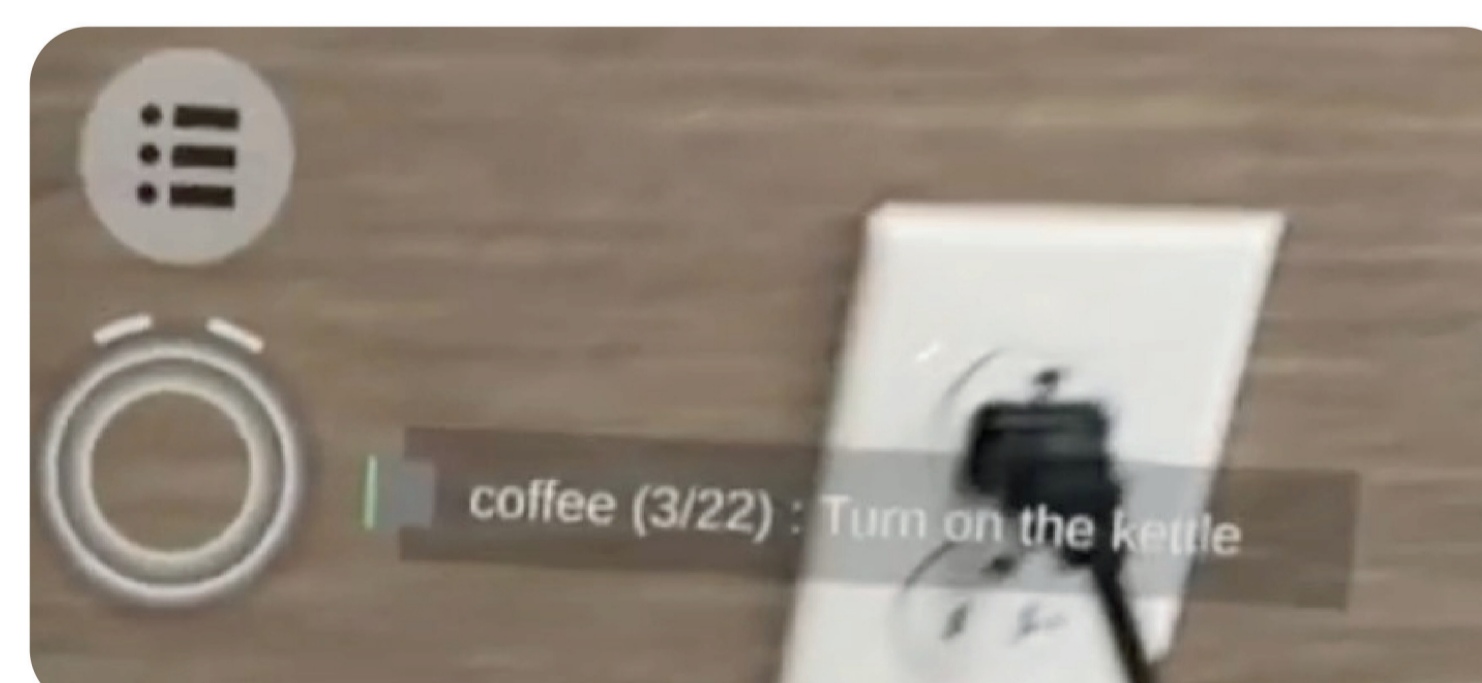
Text-based task instructions and progress are shown on demand if the user looks at the agent.

The agent's body extends into a halo to notify the user about newly emerging information, such as warnings or cautions to prevent errors.

The user can ask for task information, such as further instructions or material, if the current step description does not suffice.

The agent remains within the field of view of the display.

The agent avoids staying inside physical structures, such as walls or furniture, so the user can reach out to the agent any time (e.g., to manually adjust the position).



References

- [1] M. Gilles and E. Bevacqua. "A review of virtual assistants' characteristics: recommendations for designing an optimal human-machine cooperation." *ASME.J. Comput. Inf. Sci. Eng.* 22(5):050904, 03 2022. doi: 10.1115/1.4053369
- [2] X. Xu, A. Yu, T. R. Jonker, K. Todi, F. Lu, X. Qian, J. M. Evangelista, Belo, T. Wang, M. Li, A. Mun, T.-Y. Wu, J. Shen, T. Zhang, N. Kokhlikyan, F. Wang, P. Sorenson, S. Kim, and H. Benko. "XAIR: A framework of explainable AI in augmented reality." In *Proc. CHI*, pp.1-30, 2023. doi: 10.1145/3544548.3581500
- [3] B. Bell and S. Feiner. "Representing and processing screen space in augmented reality." In M. Haller, M. Billinghurst, and B. Thomas, eds., *Emerging technologies of augmented reality: Interfaces and design*, pp. 110-137. IGI Global, 2007. doi: 10.4018/978-1-59904-066-0.ch006

Acknowledgments

This work was supported in part by DARPA Award HR001122C0007 to Kitware, Inc. We thank Zhou Yu, Max Chen, and Derek Ahmed for developing the natural language processing components to enable a spoken dialogue between the virtual agent and the user.