# Intermediated Reality with an AI 3D Printed Character

Llogari Casas
3FINERY LTD
Edinburgh Napier University
United Kingdom
llogari@3finery.com

Kenny Mitchell
3FINERY LTD
Edinburgh Napier University
Roblox
United Kingdom
kenny@3finery.com

## ABSTRACT

We introduce live character conversational interactions in Intermediated Reality to bring real-world objects to life in Augmented Reality (AR) and Artificial Intelligence (AI). The AI recognizes live speech and generates short character responses, syncing the character's facial expressions with speech audio. The Intermediated Reality AR warping technique allows for a high degree of realism in the animated facial expressions and movements reusing live video optical appearance direct from the device's embedded camera. The proposed applications of Intermediated Reality with AI are exemplified through the captivating fusion of these technologies in toy interactive storytelling broadcasts and social telepresence scenarios. This innovative combination allows individuals to engage with AI characters in a natural and intuitive manner, creating new opportunities for social engagement and entertainment.

## CCS CONCEPTS

• **Computing methodologies → Mixed / Augmented Reality**.

## KEYWORDS

Augmented Reality, Intermediated Reality, Artificial Intelligence

**ACM Reference Format:**

Llogari Casas and Kenny Mitchell. 2023. Intermediated Reality with an AI 3D Printed Character. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live! (SIGGRAPH '23 Real-Time Live!), August 06-10, 2023.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3588430.3597251

## 1 INTRODUCTION

The Augmented Reality live AI character interaction showcase is the result of a groundbreaking technology that employs a photorealistic Intermediated Reality [Casas and Mitchell 2019] technique. Morphing optically captured real-time lit appearance allows us to animate live facial performance from generative AI characters (e.g. *Quin*), and critically relays processing via the *cloud* intermediated by the physical 3D printed object.

Real-time live speech recognition processes the speaker's utterances to textual token sequences, which are then used as a prompt for the AI response generation text. The generative pre-trained transformer (*GPT-3.5-turbo*) is an autoregressive language model the produces short responses in character, which are then processed into phonemes transitions that animate *Quin*'s facial rig parameterization in synchronization with his voice audio generation. The Intermediated Reality video warping technique allows for a high degree of realism in the animated facial expressions and movements reusing live video optical appearance from the device's embedded camera. Through a kind of optical computation reusing retargeted camera samples directly, our dynamic lighting demonstration highlights the effectiveness of the technology, showcasing the ability to create dynamic and realistic illumination and shadows in real-time.

The use cases proposed for Intermediated Reality with AI are also showcased in this work. The first proposed use case involves toy character episodic interactive story telling broadcast with hybrid AI. This is illustrated to enable remote groups to interact with their favorite characters in real-time. The second depicted use case involves social telepresence between humans and AI within the MagiPeeps metaverse. We enable individuals to interact with AI characters in a more natural and intuitive way, creating new opportunities for social engagement and entertainment.

## 2 INTERMEDIATED REALITY

Intermediated Reality [Casas and Mitchell 2019] is a term used to describe a Mixed Reality experience that combines elements of both Augmented Reality (AR) and Mediated Reality [Mann 1999] with the outcome of providing telepresence via remote physical objects. This technique creates a seamless integration of virtual content into the real-world by enabling real-time interaction and manipulation of virtual objects within a physical environment (see figure 1).

In Intermediated Reality, virtual objects, or characters as demonstrated herein, are superimposed onto the real world, appearing as if they exist in the same space. Participants can interact with these virtual elements and manipulate them while still maintaining a sense of presence in the real-world. Underpinning our previous works [Casas et al. 2018, 2017], we apply the concept of image retargeting to AR using *Object Deformed Retargeting* [Casas et al. 2017], in the case where we have known segmented 3D geometry of physical objects in the image and the ability to register the mesh pose in real-time with marker-less tracking [Comport et al. 2006]. Additionally, when these animated real-world objects cast shadows, we use *Shadow Retargeting* [Casas et al. 2018] to account for their deformation. Hence, our application enhances the user's imagination by presenting extended capabilities, such as speech and movement, to inanimate objects in the real-world. This integration opens up a multitude of possibilities across different fields, such as entertainment, communication and telepresence, having the power to transform how we interact with digital content, bridging the gap

**Figure 1: Depiction of the seamless transformation of our 3D printed character, *Quin*, into a living entity through the utilization of *Intermediated Reality*. *Quin* possesses real-time response capabilities enabled by Artificial Intelligence.**

between virtual and physical worlds to create engaging, immersive, and meaningful experiences.

## 3 IMMERSIVE AI INTERACTIVITY

Intermediated Reality interactivity with Artificial Intelligence (AI) involves the seamless integration of virtual content and real-time AI-driven interactions within a mixed reality experience. By integrating Large Language Models (LLMs) as part of the AI engine, such as GPT-3,3.5-turbo,4 or BART [Zhao et al. 2023], with Intermediated Reality, virtual characters become capable of engaging in dynamic and intelligent conversations with users (see figure 1). LLMs process and understand natural language inputs from users, allowing for voice commands and text-based interactions. Following that, the AI dynamically produces brief character responses, which subsequently undergo processing to transform them into animated phonemes [Yavas 2011]. Each phoneme is smoothly integrated with their immediate and previous tokens, allowing for a natural blending of visemes while controlling its blend weights acceleration to ensure physical plausibility. These visemes are then utilized to animate the facial rig parameters of the characters, ensuring precise synchronization with the generation of voice audio. This empowers users to engage in interactive and lifelike conversations with virtual characters, while also receiving information tailored to their specific interests, enabling natural interactions between users and virtual elements.

## 4 IMPLEMENTATION

Our character, *Quin*, was initially created through traditional 3D modeling software and then brought to life using full-color 3D printing in sandstone. Further, this 3D asset was subsequently animated using blend-shapes to represent each of the 44 English phonemes and a point-cloud was generated to facilitate marker-less object tracking in Augmented Reality. The framework has been natively implemented for iOS devices using *Swift* and *Metal* programming languages, harnessing the capabilities of [ARKit 2022].

In our implementation, the *SFSpeechRecognizer* library is utilized for speech recognition, which transcribes the text and sends it to a *Large Language Model* API for processing. The resulting response is then directed to the *AVSpeechSynthesizer* library for speech generation while procedurally animating the character expressions based on the AI generated response. The speech recognition and synthesis libraries run on the device using hardware-accelerated methods, while the LLM relies on cloud for responses.

## 5 APPLICATIONS

The integration of LLMs with IR technology offers a wide spectrum of applications spanning a variety of industries presenting immersive and captivating interactive experiences across the board.

In the realm of virtual assistants, LLMs enable intelligent conversational agents that can understand and respond to user queries, providing personalized recommendations and facilitating natural language conversations. This enhances customer support and information retrieval. Interactive storytelling benefits from IR and LLMs as virtual characters can dynamically respond to user inputs, creating personalized and branching narratives that captivate users and offer unique storylines. Education is another domain where IR with LLMs finds application. LLMs can assist in creating interactive simulations and virtual mentors, offering personalized and engaging learning experiences. Students can ask questions, receive explanations, and interact with virtual characters to deepen their understanding. Language learning becomes more interactive with IR and LLMs. Virtual language partners powered by LLMs provide conversational practice, real-time feedback, and simulations of immersive language immersion environments. Further, training scenarios benefit from IR and LLMs by offering realistic simulations with virtual characters that respond to trainee actions and provide real-time guidance. These applications demonstrate the power of combining IR and LLMs to create enriching user experiences and pushing the boundaries of virtual interactions.

## ACKNOWLEDGMENTS

The authors express their gratitude to Samantha Hannah and Alex Cullinane-Carrasco for their contributions in developing and animating the 3D assets utilized in our demonstration. This project has received funding from InnovateUK's grant agreement No. 68475 and from European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017779.

## REFERENCES

ARKit. 2022. Apple ARKit. https://developer.apple.com/augmented-reality/arkit/

Llogari Casas, Matthias Fauconneau, Maggie Kosek, Kieran Mclister, and Kenny Mitchell. 2018. Image Based Proximate Shadow Retargeting. In *Computer Graphics and Visual Computing (CGVC)*, Gary K. L. Tam and Franck Vidal (Eds.). The Eurographics Association, Swansea, UK. https://doi.org/10.2312/cgvc.20181206

Llogari Casas, Maggie Kosek, and Kenny Mitchell. 2017. Props Alive: A Framework for Augmented Reality Stop Motion Animation. In *2017 IEEE 10th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS)*. 1–4. https://doi.org/10.1109/SEARIS41720.2017.9183487

Llogari Casas and Kenny Mitchell. 2019. Intermediated Reality: A Framework for Communication Through Tele-Puppetry. *Frontiers in Robotics and AI* 6 (2019). https://doi.org/10.3389/frobt.2019.00060

A.I. Comport, E. Marchand, M. Pressigout, and F. Chaumette. 2006. Real-time markerless tracking for augmented reality: the virtual visual servoing framework. *IEEE Transactions on Visualization and Computer Graphics* 12, 4 (2006), 615–628. https://doi.org/10.1109/TVCG.2006.78

Steve Mann. 1999. Mediated Reality. *Linux J.* 1999, 59es (mar 1999), 5–es.

Mehmet S. Yavas. 2011. *Phonetics.* John Wiley Sons Ltd, New York, NY, USA, Chapter 1, 1–29. https://doi.org/10.1002/9781444392623.ch1

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:cs.CL/2303.18223