

Mixed Reality Communication System for Procedural Tasks

Manuel Rebol*
mrebol@american.edu
American University
Washington, USA

Erin Maria Horan
ehoran@american.edu
American University
Washington, USA

Yasser Ajabnoor
yajabnoor@mfa.gwu.edu
George Washington University
Washington, USA

Claudia Ranniger
cranniger@mfa.gwu.edu
George Washington University
Washington, USA

Adam Rutenberg
arutenberg@mfa.gwu.edu
George Washington University
Washington, USA

Safinaz M Alshiakh
salshiakh@mfa.gwu.edu
George Washington University
Washington, USA

Colton Hood
chood@mfa.gwu.edu
George Washington University
Washington, USA

Neal Sikka
nsikka@mfa.gwu.edu
George Washington University
Washington, USA

Krzysztof Pietroszek
pietrosz@american.edu
American University
Washington, USA

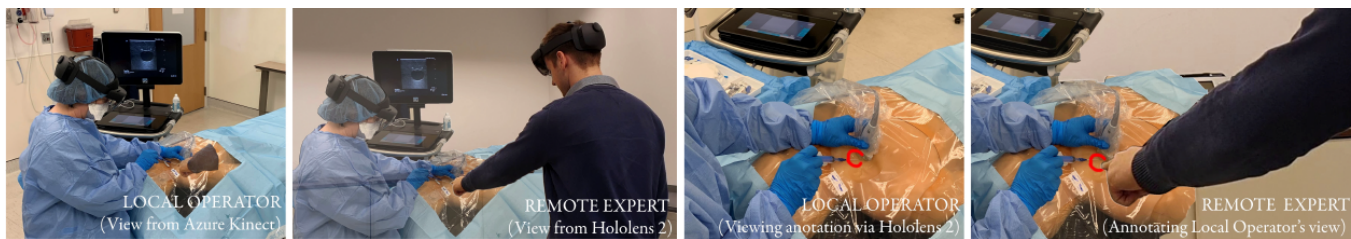


Figure 1: Volumetric communication system: 1. Local operator sees a virtual hand of the remote expert while remote expert sees a volumetric representation of the local operator and the patient (left). 2. Local operator and remote expert can create, view and change “in situ” annotations

ABSTRACT

We design a volumetric communication system for remote assistance of procedural medical tasks. The system allows a remote expert to spatially guide a local operator using a real-time volumetric representation of the patient. Guidance is provided by voice, virtual hand metaphor, and annotations performed *in situ*. We include the feedback we received from the medical professionals and early NASA TLX [5] data on the cognitive load of the system.

CCS CONCEPTS

• **Computing methodologies** → **Mixed / augmented reality; Volumetric models.**

KEYWORDS

Remote Procedure, Augmented Reality, Virtual Presence.

*Also with Graz University of Technology.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
AVI 2022, June 6–10, 2022, Frascati, Rome, Italy
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9719-3/22/06.
<https://doi.org/10.1145/3531073.3534497>

ACM Reference Format:

Manuel Rebol, Claudia Ranniger, Colton Hood, Erin Maria Horan, Adam Rutenberg, Neal Sikka, Yasser Ajabnoor, Safinaz M Alshiakh, and Krzysztof Pietroszek. 2022. Mixed Reality Communication System for Procedural Tasks. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces (AVI 2022)*, June 6–10, 2022, Frascati, Rome, Italy. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3531073.3534497>

1 INTRODUCTION

Remote collaboration is playing an increasingly important role in people’s everyday lives with video communication has become a *de facto* standard for remote collaboration. Yet, in some applications, lack of spacial information limited field of view, lack of context, and a limited transmission of non-verbal cues and body language, video communication is insufficient [9]. One example where traditional video communication is not sufficient is remote assistance in procedural medical tasks [3, 4, 6, 11].

We describe the design and implementation of a communication system that can be used for remote assistance in procedural medical tasks. To increase the accessibility and lower the costs, we develop this system using off-the-shelf hardware: Hololens 2 [2] and Kinect Azure[1].

2 SYSTEM DESIGN AND IMPLEMENTATION

Our system supports audiovisual, spatially-enhanced communication and interaction between two parties: the local operator and

the remote expert. The local operator works on a procedural task while being assisted by a remote expert, who provides spatial and auditory guidance.

Both users are equipped with a lightweight, standalone, high-resolution mixed-reality head-mounted display. For our prototyping, we choose the Hololens 2, but in principle, any mixed-reality headset capable of freehand interaction could be used.

Additionally, the local operator's environment is equipped with a volumetric capture camera that continuously feeds high-fidelity spatial data to the remote expert. We use the volumetric capture of the low-cost time-of-flight depth camera Kinect v4 Azure with a separate processing unit. In our setup, the volumetric capture is performed from the position where a supervising operator would have positioned themselves if they were co-located with the local operator.

To provide the remote expert with spatial information on the local scene, our system continuously “scans” the object of interest tracks the position of the hands and tools used by the local operator, and presents this information as a real-time volumetric representation. Importantly, the spatial sound of the local operator's environment is also communicated to the remote expert, in an attempt to create the illusion of co-location for the remote expert.

Using the spatially-rendered view of the local operator's object of interest and its context, the remote expert visually guides the local operator in performing a procedure. The expert uses speech, pointing, and annotating topology of the object of interest.

An example of the view of the expert is shown in Figure 1. As illustrated, the remote expert can annotate the view and that annotation is visible to the local operator. The view of the local operator includes the virtual hand of the remote operator and the remote operator's annotation presented on the physical object of interest, e.g. the patient. The local operator can annotate any physical object of interest and these annotations are visible to the expert, providing two-way annotation-based communication between the local and the remote party.

At the local site, the scene is captured with the Kinect Azure (Kinect v4) volumetric camera. The color information is retrieved by the RGB camera and the depth by the time-of-flight sensor. We process both data streams using the Azure Kinect SDK. Once the streams are processed, they are forwarded to the Mixed Reality WebRTC [10] libraries that handle the peer-to-peer connection between the remote and local operator. The camera sends both depth and color data in real time over network to the remote expert.

The WebRTC client running on the head-mounted display of the remote and the local operator receives the depth streams. Additionally, the remote expert receives the color stream. For the remote view, we assemble the depth and color information and render a 3D mesh. As a result, the local scene is visualized with high spatial fidelity on the Hololens 2 of the remote expert.

3 EVALUATION

We are currently performing a comprehensive evaluation of the system with medical professionals and medical students at the Clinical Learning & Simulation Skills Center of George Washington University. In the evaluation reported here, we compared video communication currently used for a remote procedure with our

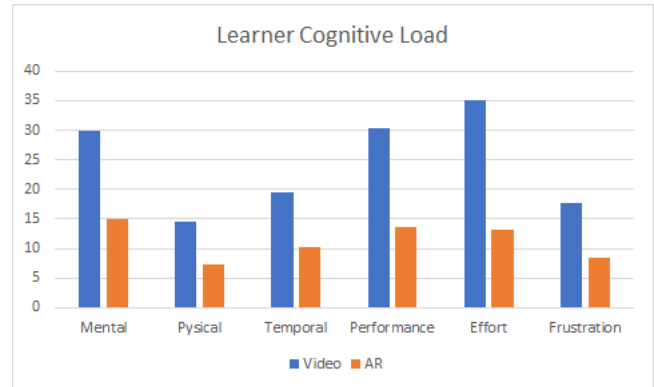


Figure 2: NASA TLX. We compare the learner's cognitive load using the NASA TLX instrument. The bars indicate the mean TLX score in each cognitive load category for both video and AR training.

prototype mixed-reality volumetric communication for the task of assisting in the placement of a central venous catheter. This life-saving, but risky, medical procedure was performed on an Internal Jugular Central Line Ultrasound Manikin, not a real patient. Our study design was between subjects. We had six expert users guiding 16 local operator participants who have never performed the procedure before, placed in two groups: control (the video assistance, 8 participants), and intervention (the volumetric communication, 8 participants). The local operators were drawn from the students and medical residents of the medical school.

We used the NASA TLX [5] survey as a quantitative cognitive load instrument. Our null hypothesis was that the mental load, effort, and fatigue are not significantly reduced for our volumetric system. We expect this null hypothesis to be rejected due to the availability of spatial information and non-verbal cues absent in video communication.

We used an independent sample one-sided t-test to test our null hypothesis. Our null hypothesis assuming equality of means was rejected for cognitive load, with respect to temporal demand, performance, and effort being significant ($p < 0.05$). No significance was found for the other TLX components mental, physical, and frustration. As seen in Figure 2, there seems to be a trend toward significance in mental, physical, and frustration, which suggests a larger sample is required to reject the null hypothesis for those components of NASA TLX. The results suggest that the cognitive load of the learner using MR is lower than using video communication.

4 CONCLUSION

We presented a real-time volumetric communication system for medical applications. We evaluated our prototype on a central venous catheter placement procedure. The cognitive load measurements of learners using our volumetric communication system were lower than using baseline video communication. In future work, we plan to evaluate remote spatial communication training in other procedures. Moreover, we want to augment automated avatar representations [7, 8] for recorded instructions.

ACKNOWLEDGMENTS

We would like to thank medical students and residents for participating in the experiment. The work is supported by National Science Foundation grant no. 2026505 and 2026568.

REFERENCES

- [1] 2022. Microsoft Azure Kinect. <https://azure.microsoft.com/en-us/services/kinect-dk/#overview> .
- [2] 2022. Microsoft HoloLens 2. <https://www.microsoft.com/en-us/hololens/buy> .
- [3] Danilo Gasques, Janet G Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, et al. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telementoring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. .
- [4] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. 2012. TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 619–622. .
- [5] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage Publications Sage CA: Los Angeles, CA, 904–908. .
- [6] D. Mourtzis, V. Zogopoulos, and E. Vlachou. 2017. Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry. *Procedia CIRP* (2017), 46 – 51. .
- [7] Manuel Rebol, Christian Güti, and Krzysztof Pietroszek. 2021. Passing a non-verbal turing test: Evaluating gesture animations generated from speech. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 573–581. .
- [8] Manuel Rebol, Christian Güti, and Krzysztof Pietroszek. 2021. *Real-Time Gesture Animation Generation from Speech for Virtual Human Interaction*. Association for Computing Machinery. .
- [9] Manuel Rebol, Colton Hood, Claudia Ranniger, Adam Rutenberg, Neal Sikka, Erin Maria Horan, Christian Güti, and Krzysztof Pietroszek. 2021. Remote Assistance with Mixed Reality for Procedural Tasks. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 653–654. <https://doi.org/10.1109/VRW52623.2021.00209> .
- [10] G. Suciu, S. Stefanescu, C. Beceanu, and M. Ceaparu. 2020. WebRTC role in real-time communication and video conferencing. In *2020 Global Internet of Things Summit (GloTS)*. 1–6. .
- [11] S. Wang, M. Parsons, Jordan Stone-McLean, P. Rogers, S. Boyd, K. Hoover, O. Pastor, Minglun Gong, and A. Smith. 2017. Augmented Reality as a Telemedicine Platform for Remote Procedural Training. *Sensors (Basel, Switzerland)* 17 (2017). .