

Preference in Voice Commands and Gesture Controls With Hands-Free Augmented Reality With Novel Users

Mikko Korhikoski , Paula Alavesä , and Panos Kostakos , University of Oulu, 90570, Oulu, Finland

Hands-free augmented reality (HAR) has the potential of one day becoming our gateway to the Metaverse. However, the devices are still out of reach for most consumers mainly due to their price. The issues with novel interaction techniques can also present a challenge when interacting with holographic content in 3D space. In this article, we describe a mixed-method user study (N=34, inexperienced HAR users) with an augmented reality (AR) map application and HoloLens 2, where the target is to compare the usability, user experience (UX), and preferences with two of the most common HAR interaction techniques: gestures and voice commands. Our findings show that there may not be a significant difference in the overall UX between these controls when using an AR map. However, there are individual variability and differences in how novel users prefer to interact with holographic content. Gesture controls were preferred for their playfulness and voice commands for their efficiency. This suggests that developers should focus on different interaction techniques for different contexts from gaming to serious collaborative work.

Hands-free augmented reality (HAR) remains somewhat rare due to the high price and availability of smart glasses and head-mounted displays (HMDs) with the feature. Most people have no prior experience with HAR; however, as these devices become more lightweight and affordable, they have the potential to even replace our mobile phones and become the default technology for accessing the Metaverse. We are now at the stage of technology adoption, where problems influencing usability and user acceptance should be researched and solutions found.¹

Mobile augmented reality (MAR) is often used in applications such as chat filters or barcode readers. Many people are already unknowingly using augmented reality (AR). Some AR applications like Pokémon Go have been successful, but these

examples are only a few. A major contributor to the success of the aforementioned applications is possibly the platform: the mobile phone. For MAR applications, the potential customer base is the entire world. While these MAR applications have their place, the bigger potential of AR lies in HAR. The current roadblock for HAR is the cost of devices such as Meta Quest Pro and HoloLens 2, making them unappealing and unattainable to the broader audience. In the foreseeable future, roadblocks may no longer exist and HAR devices can achieve similar mobility and ubiquity as mobile phones. It can also be expected that AR functionality will complement immersive VR allowing even more scoping access to digital realms such as the Metaverse.^{1,2,3,4}

In this article, the Augmented Reality Tactical Sandbox (ARTS), a map application developed for HoloLens 2 HMD, is used as a probe in a mixed-method user study (N = 34). The goal of this work is to scope the current and future issues influencing user acceptance of HAR devices. We approach this issue by comparing the two most common interaction techniques used in HAR: voice commands and gestures to evaluate user preferences.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>
Digital Object Identifier 10.1109/MPRV.2024.3364541
Date of publication 4 March 2024; date of current version 30 May 2024.

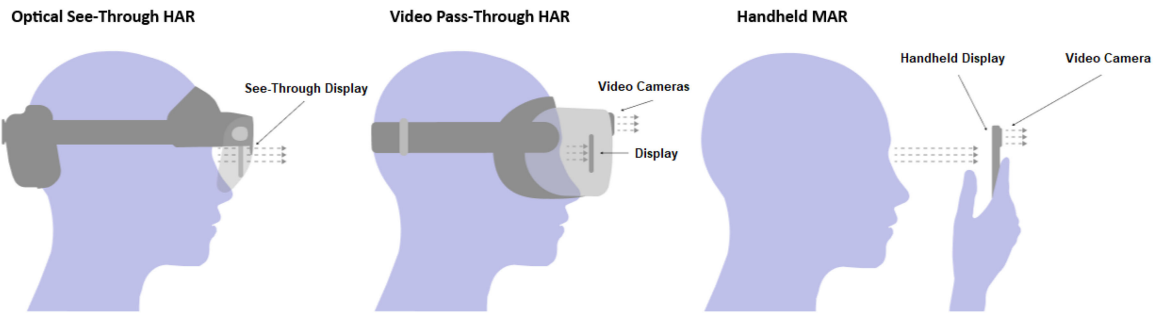


FIGURE 1. Examples of currently available AR technologies. Left to right: optical see-through (Microsoft HoloLens 2) HAR, video pass-through (Meta Quest Pro) HAR, and MAR. The two leftmost are mainly used for hands-free while MAR requires the use of hands for holding the device and is therefore intrinsically different.⁸

CURRENT WAYS TO ACCESS AR

There are many technologies used to access AR content: mobile-based, projection-based, see-through-based, or even PC-based. All of these technologies provide different multimodal ways to interact with AR content.⁵ AR can also be divided into other subcategories according to how it is accessed. The most common type of AR is MAR, where a hand-held mobile phone is used and interfaced with a 2D touch-screen. A less common type of AR is HAR, where the device is worn on the user's head like glasses. This leaves hands free for interacting with the holographic content. The ability to interact with digital items in 3D space is such a groundbreaking addition to AR that, to highlight this difference in this article, we have chosen to call these technologies hands-free AR, HAR. The interaction techniques in HAR vary between devices, but the main techniques are gestures, voice commands, or handheld controllers. Using gestures does not require the user to have anything in their hands, so they are, in essence, hands-free, and the interactables are situated around the user in a three-dimensional (3D) space. In contrast, with MAR, they appear to be in 3D space, but interactions are mediated by a 2D surface.^{6,7} The used gestures mimic familiar hand gestures such as pinching, pointing, and grabbing. Immersion and usability between MAR and HAR can differ substantially.²

There are HAR devices that use see-through displays to visualize AR content, while the other technique is to use video pass-through to display the AR content on a display close to the user's eyes. Figure 1 illustrates the difference between MAR and HAR. Handheld controllers can be used for HAR, however, for more pervasive daily use voice commands, gestures,⁹ or more ubiquitous controllers such as smart rings show more promise.

UX, usability, and user preferences with different interaction techniques are widely researched even with AR, but the main focus of the current research is on interaction mediated by 2D surfaces spanning from the mobile phone display to AR projected on car windshields.^{10,11,12}

Interactions in HAR

There are three common techniques for interacting in HAR, and these techniques differ in usability and performance. Research has shown that the combination of both voice commands and gestures is the fastest, with only a slight difference to using voice commands only, while gestures are found to be the slowest.¹³ However, many nuances and details are still inconclusive on what is the most efficient and intuitive control technique for HAR. Some users still prefer to use gestures instead of voice in certain tasks, even when the performance is worse.¹⁴ Some state voice-only or voice combined with gestures feels more natural than only gestures.^{9,13} In some situations, if hands are already in use, voice commands can enhance the operation speed and efficiency.¹⁴ To summarize, in comparison to hand-held controllers,¹⁵ gestures can have better performance in some tasks and they can be easier to learn, but on the other hand, using them can be physically tiring.

HoloLens 2 HAR Device

The very first HMD developed by Sutherland in 1968 was an AR HMD, however, it took decades for this technology to become miniaturized into marketable products.² In 2016, Microsoft released its first generation HAR device: HoloLens, that used optical see-through (Figure 1) to display the virtual content. Three years later, in 2019, the HoloLens 2 was released. It

used the same kind of display technology as its predecessor and featured improved design and more powerful hardware.

HoloLens 2 uses the Articulated HAnd Tracking (AHAT) depth camera in two different modes. For capturing hand movements to obtain hand-tracking data, the high-frequency (45fps) near-depth sensing mode is used. The low-frequency (low-throw) (1-5fps) far-depth sensing mode is used for spatial mapping. The obtained mapping data can then be used to place holographic content into surrounding space. Visible Light Environment Tracking cameras (grey-scale) are used for head-tracking and map building while an infrared (IR) reflectivity stream is used to compute depth. The device also includes an inertial measurement unit for detecting rotations and orientation.¹⁶ The precision of the hand tracking of the device is not known, though aspects influencing the tracking precision have been researched.^{6,7,17}

Maps in AR Versus Traditional Monitor

A study¹⁸ compared a monitor map application, a Microsoft HoloLens version, and the Augmented Reality Sandtable (ARES), which is a mix of a real-world sand table terrain and a down-projected 2D map. In a situational judgment task, where the participants needed to rank different route alternatives, the AR application allowed the users to finish the task significantly quicker than with the regular 2D application. The participants were also significantly more engaged while using the AR version. We continue where this study finished and aim to investigate which AR interaction technique users prefer in AR.

AUGMENTED REALITY TACTICAL SANDBOX

ARTS has been developed as a part of a bigger system for border surveillance. The requirements were defined with a user-centered design approach. The purpose of this application is to provide a 3D AR map for better situational awareness in the command and control center and in the field. The application displays a 3D map on the HoloLens 2 HMD (Figure 1). The application was developed on top of Microsoft's Bing Maps, a commercial product with high quality requirements. The basic map functionalities therefore inherit many best practices for designing digital maps. Bing Maps also allowed the addition of custom content and analytics for research purposes.

When ARTS starts up, the holographic map (Figure 3) is placed roughly 1.5 m below the user's gaze point, placing it below where the user is looking rather

than the point where their eyes are. The user can always resituate the map. The map can also be resized or rotated. Map placement can be done via voice commands or gesture controls while resizing and rotating must be done with gestures. During the user test, the users accessed these functions to better align the map with a physical table used for reference.

The users can also move freely into a different area on the map using voice commands or gestures. Using the voice command "pan" followed by "north/south/west/east," the users can move the map slightly to the desired direction. With the gesture "air pinch," the users can "grab" the map and then drag it in any direction.

ARTS contains two "quick travel" buttons and four voice commands that allow the user to change the position of the map to preset coordinates. In the experiment, the users used this function with voice commands and gestures. The gesture control is simply to press a button for the function. Voice commands are either "Return (to) Home/Greece" or "Go to New York." Zoom functions are also available with buttons or voice commands. The zoom function gives the user the ability to view any location at any height. ARTS also allows the users to hide assets (boats, drones, vehicles) from the map. This can also be done with either gestures or voice commands.

For gestures, the application has a hovering menu bar [see Figure 3(b)/(e)]. In addition, users can "air pinch" or "physically" pinch the virtual object for direct interactions. The menu bar can be hidden or shown with a voice command and pinned down anywhere. By default, the menu will follow the user as they move.

METHOD

A mixed-method task-based user study with a qualitative priori was conducted between November 2022 and March 2023 in a controlled lab environment. The interaction technique was the independent variable in this study. The goal was to compare the usability, UX, and preferences in voice command and gesture controls of participants ($N = 34$) who had no prior experience with AR smart glasses such as HoloLens 2.

Study Procedure

The study procedure (see Figure 2) began with the participant reading and signing consent forms. As most people in the general public have no experience with HAR,¹ we made this a prerequisite in our study. We only had to exclude one participant due to their prior HAR experience. Eleven out of 34 users had prior experience with MAR, but this was expected.

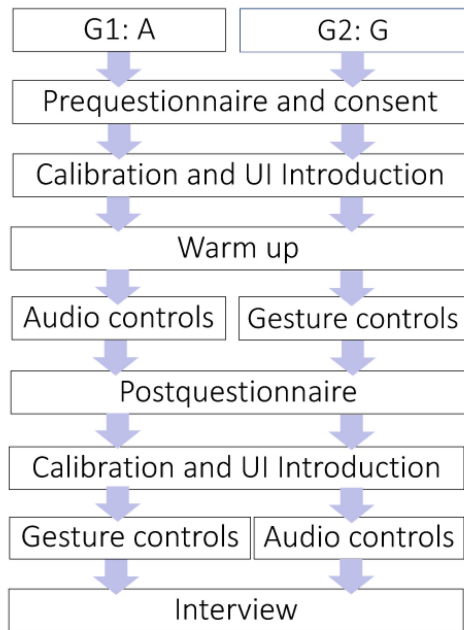


FIGURE 2. Task sequence for voice command (G1-A) and gesture (G2-G) groups.

Next, the participants were introduced to the HoloLens 2 HMD followed by an introduction to basic AR concepts, the testing procedure, and ARTS. The participants were divided into two groups: one group would use voice commands first and the other would use gestures first (Figure 2).

For the group using voice commands first (G1-A in Figure 2), the first task (Task 1: Warm up in Figure 2) would consist of running the eye calibration for the HoloLens 2 and a quick introduction to the basic user interface (UI) controls. For the group using gestures first (G2-G in Figure 2), this warm-up task was the same, but with a tutorial for the “air pinch” gesture. This tutorial was done using an interactable floating 3D football [see Figure 3(a)]. When they felt confident about using the “air pinch” gesture, the warm-up task would end. This was done to exclude the influence¹⁹ of difficulties in “air pinch” for those who take longer to learn the gesture. The G1-A group would perform the “air pinch” tutorial at the beginning of task 3 (Task 3: Gesture in Figure 2).

The tasks for both groups were exactly the same but done using a different interaction technique; voice commands or gestures. For the group using voice commands, there was a quick voice check at the beginning. The participant disabled and enabled the gesture menu with voice commands, and if everything worked they proceeded with the task. In the gesture group, participants observed how the menu followed their head movements and learned how to disable its movement so it would not follow them. When the turns changed to use a different interaction technique, the same procedure for checks was followed. The participants were shown a list of tasks in view to do with ARTS during the experiments. Each participant was encouraged to do their tasks independently and only ask the researchers if there was a clear problem.

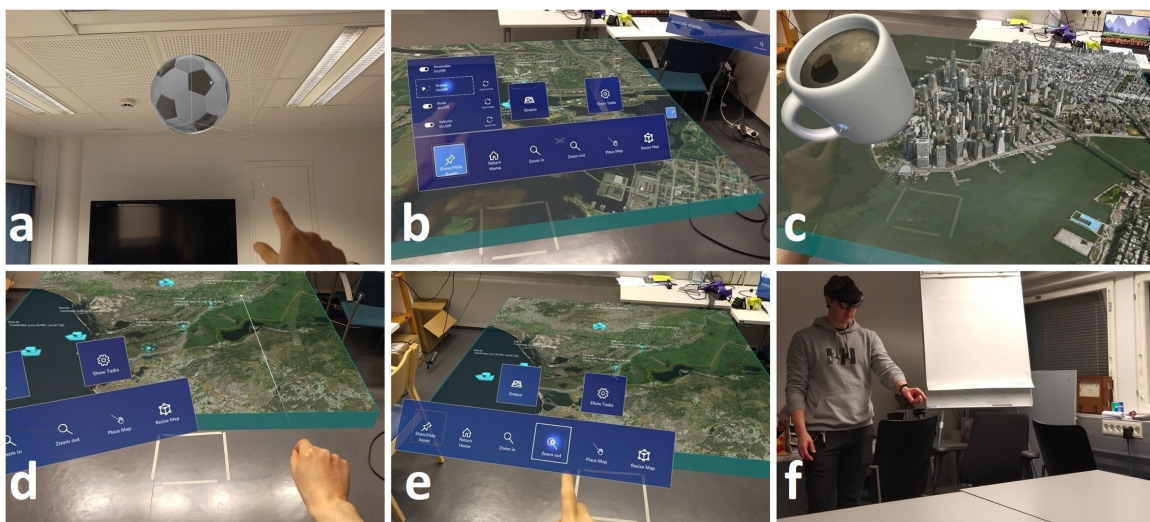


FIGURE 3. (a) Virtual football used for gesture training. (b) Floating user menu showing all buttons. (c) Final phase of testing at New York + coffee cup interaction (d) Moving the map with the air pinch gesture (e) Menu button press interaction. (f) Study participant using the map during the experiment.

The tasks (see Figure 3) were identical for both groups, the only difference was the interaction technique. The G1-A group used voice commands first (Task 2: Audio, Figure 2.), then gestures second (Gesture controls, Figure 2), whereas the sequence was opposite for the G2-G group. After answering the post-questionnaire, the participants went through the same tasks but with the interaction technique they had not yet experienced. After completing the main tasks there was a bonus task, where the participants had the option, though not explicitly suggested, to use both interaction techniques at the same time. They were also asked to try out the hidden coffee cup voice command. The bonus task was a way to reward the users, but also allowed us to observe what technique they chose to use. The task sequence was as follows:

Tasks.

- 1) Use the zoom-in/out functions.
- 2) Pan the map.
- 3) Move the map to "home."
- 4) Use the show/hide functions for different assets.
- 5) Move the map to "Greece."

Additional assignment, which allowed but did not force mixed interaction techniques, at the end of tasks.

- 1) Move the map to New York.
- 2) Find the Statue of Liberty.
- 3) Make the coffee cup appear and interact with it.
- 4) Use the map any way you want (optional).

Material

The collected material consisted of informed consent, a prequestionnaire for demographics, a postquestionnaire, and a brief transcribed interview. The postquestionnaire included the system usability scale²⁰ (SUS) questionnaire and the user experience questionnaire (UEQ).²¹ It was issued after the users only had experience with one interaction technique. At the very end of the experiment, following the participant's experience with both interaction techniques, we conducted a semistructured interview. The interview targeted their favorite interaction technique, preference of use, and general feedback. The average length for an interview was 3 min 57 s.

Participants

We recruited 34 participants (M = 23, F = 11) via university mailing lists and an online recruitment system that prescreened for participants with no prior

experience using HAR devices. Most participants were students with a major in computer science and engineering or some other field of engineering, however, there were also archaeology and logopedics majors. The age range of the participants was 21–48 (M = 27.2, SD = 5.2). The majority (20) of the participants were of Finnish nationality.

RESULTS

Based on the SUS and UEQ, issued after the users had a chance to try either the voice command or the gesture interaction technique, there were no significant differences in basic usability and UX between these techniques. However, in the following section, we summarize the analysis methods and the small gender-related outlier observations. The questionnaire results are between-subjects as they were collected before the users had a chance to try both modes of interaction.

System Usability Scale

The SUS questionnaire results were analyzed using the independent-samples T test. There were no significant differences in the overall SUS score, but one significant difference was found between the groups in one of the questions. The difference was in question 6 (N = 32, mean G1-A = 1.733, mean G2-G = 2.588, $p = 0.014$) on the inconsistency in the system. The group that used gestures (G2-G) evaluated the system as being (significantly) more inconsistent than the G1-A group, which used voice commands. The overall SUS score showed no differences apart from one question, which showed a small gender difference between men and women in G1-A in question 7 (N = 15, M = 11, F = 4, mean rank M = 6.36, mean rank F = 12.50, $p = 0.012$). Here, the women in the group evaluated the system as being significantly quicker to learn than men.

User Experience Questionnaire

The UEQ (scale from (-3) to 3) was analyzed using both one-way ANOVA and the Kruskal-Wallis H test. We found no significant differences in UX between the groups for the six main categories [$p = (0.212, 0.295, 0.356, 0.442, 0.937, 0.940)$]. However, between genders in the G1-A group, there was a significant difference in attractiveness (N = 16, mean M = 1.79, mean F = 1.18, $p = 0.009$) of the application, with men rating it higher than women. In the G2-G group, the women rated efficiency higher than men (N = 13, mean M = 0.67, mean F = 2.25, $p < 0.000$). Differences between genders were analyzed using one-way ANOVA.

Qualitative Findings

The qualitative findings are based on the semistructured interviews conducted after the users had a chance to try both voice command and gesture interactions. Therefore, while these observations are within-subjects, half of the users tried voice commands first and half tried gestures first.

User Preferences

The participants had initial preferences for either voice commands (17), gestures (12), or mixed (5) controls as their first choice when asked about what interaction technique they found most comfortable. When enquired about what interaction technique they would use in practice, the preferences shifted in favor of mixed-use (14) followed by gestures (11) and voice commands (9). Interestingly, none of those who initially preferred gestures chose voice commands afterward, while three users shifted from voice commands to favoring gestures in practice. Agreement testing was conducted between two observers with a sample of 100 text excerpts. For recognizing whether the user was talking about the experience in general or a specific interaction, there was a very high agreement of 96% with $Kappa = 0.937$. In recognizing categories with the used interaction technique the agreement was 87% with $Kappa = 0.839$. There was slight disagreement in the interpretation of intuitiveness under the Experience category. Also, observers disagreed on four occasions about learning requirements apposed to having to remember voice commands or UI layout due to their similarity. These were also in separate categories. The thematic analysis²² of the further interviews revealed potential reasons for the preferences, which can be divided into the following three main categories:

Reliability

The users found voice commands to be faster, more reliable, and more efficient. To quote a user: "...I had choose separate things for showing and hiding the objects. I had to press the button and select the object, but with voice command it was just simple as saying the one voice command."

The issues users had with controls varied. Neither method was problem free. With gesture controls, problems arose from the technology. The most common issues were unreactivity (i.e., inertness) and overreactivity. Some users struggled especially with the pinch gesture, while some users had unintended interactivity with the content, such as the floating menu or the map sticking to their hand. To quote one user on gestures: "If I knew what type of interaction would result in what type of end-results the use would be

faster"... "Many times these balls and maps kind-off stuck to your finger. You stop the gesture, but the target follows you." With voice commands, the users' main concern was if they would remember the commands. However, both issues were seen by the users as something they would be able to overcome with practice. The pinch gesture was difficult while pointing and grabbing worked for most. Both voice command and gesture controls evoked comments about a high learning curve. For voice commands, the users felt they needed to learn the utterances, and with gestures, finding the correct hand positions was noted to require practice as well.

THERE WAS SLIGHT DISAGREEMENT IN THE INTERPRETATION OF INTUITIVENESS UNDER THE EXPERIENCE CATEGORY.

Experience

One identified category was experience, specifically, how natural and fun the users found the interactions with gestures as opposed to voice commands. The gesture controls were considered more fun, natural, intuitive, and, as an interaction method, more closely resembling how one interacts with objects in physical reality. The users were in general more impressed with gesture controls. There were also a few users whose initial preference was voice commands but still stated that they would use gestures in practice. To quote one user on the gesture-related UX: "It felt, like, more controlled. I like to play games where you can move the map view. This was kind of the same and the setup was more natural... gesture controls for instance in games, it creates this bodily... like for instance, I like playing with Wii console."

Comparison With Current Technologies

The users frequently compared the AR experience to familiar technology. This is known to occur with other novel technologies as well and might influence user acceptance in serious contexts like work. SUS scores and UEQ scales also show that no big usability or UX issues were influencing the findings. Some users compared the device to existing use cases or technology such as their PCs or smartphones, to quote a female participant from G1-A: "You cannot yet watch Netflix with it, since the colors are not like on a laptop screen..."

Participants were also enquired if they found the technology mature enough for daily use and if they would use the device themselves. 18 people answered yes, 15 said no, and one replied maybe.

Limitations

This study has a small sample size. The differences between gesture and voice controls were not statistically significant. However, the qualitative analysis gives us more insight into how the users experienced the controls and the use of HAR. For qualitative analysis, the sample size is adequate,²³ with the only issue being the brevity of the interviews due to the need to keep the study's time limit under an hour per user. We mainly recruited participants from the university campus, and many of them reported an interest in the technology. Due to this potential sampling bias, one must be careful when generalizing the overall UX or SUS scores. Since none of the participants had any prior experience with HAR, we can assume that the results from the qualitative analysis of the user preferences have high generalizability. Although, there is high interpretability due to the qualitative method. Agreement testing, though, suggests that the interpretability of the results is small. There is a chance that the perceived efficiency of voice command interaction is due to technology limitations and lack of experience. However, we aimed to mitigate this issue by having calibration, UI introduction, and a warm-up added to the study procedure.

DISCUSSION

We do not know yet if being able to use AR hands-free is the needed "killer feature" for the technology. The HAR devices can display information to the users as they walk the streets, but the possibility to interact with the environment is not yet there for a full Metaverse experience. In such a scenario, for instance, one could buy something simply by interacting with an advertisement they see, and then have Amazon deliver the product to their home a few days later.

Considering current smart glasses such as HoloLens 2, as interfaces to the Metaverse, the threshold question is: how naturally and intuitively can people, especially novel users, interact with the holographic content? Early smart glasses, like Google Glass, suffered from social rejection and privacy issues.²⁴ Even when the technology becomes more miniaturized and affordable, the public might feel reluctant to adopt HAR due to the high learning curve of available controls and reliance on phones and tablets.²⁵ In this study, the users compared the HoloLens 2 to the

technologies they are more familiar with and they currently use to interact with digital maps.

For high user acceptance of new technology such as HAR, more than just feasibility tests are needed. An important goal, when unlocking the potential of this technology, is to understand the users' needs and preferences beyond usability and basic UX. Our findings show that there are individual differences in the user preference for interaction techniques despite there not being statistically significant differences in usability or UX of the two most common interaction techniques. We noticed through the interviews and observations that individual preferences can vary and are based on how reliable the users find a specific technique. Some users complained about accidental interactions with gestures, and some had a high learning curve. While voice commands were found to be efficient, it was not perceived as exciting. However, voice commands provided a reliable option for those who had a frustrating initial experience with gestures. Still, gestures were found to be more intuitive, as many users wondered if they would be able to recollect all needed voice commands. We expected there to be a stronger novelty effect from the gestures considering that this interaction technique is more novel and voice commands have been around for decades on many consumer devices. However, the users found the gesture controls more natural and compared using them to how one interacts with objects in the physical world.

Voice commands were the initial preference for many of our test users. However, those who had no issues with gestures preferred them. Hand-tracking still seems to be an issue with gestures, but this is due to limitations in the technology and might be mitigated in the near future. Both interaction techniques had a learning curve, but once the hand-tracking issues are ironed out, voice commands will still retain the problem of users worrying about memorizing the phrases.

Design Recommendations

What HAR-type smart glasses provide is a direct way of interacting with holographic objects as one would in the physical world. Our findings show that the users consider this type of interaction exciting and intuitive. Therefore, we suggest that especially when designing for games and the context of leisure, gesture interactions should be prioritized due to the "excitement" effect and making the player feel more involved. The popularity of exergames, fitness games, and motion controllers is a good indicator that people like moving

around and using their hands while playing games. These recommendations are not exhaustive, and of course, there may be application-specific considerations that defy the context.

For serious contexts, such as work, where efficiency is the main factor, voice commands should be favored. However, one must be careful not to rely solely on voice commands since issues will arise when working in loud environments. It is recommended to have gestures as a backup. For situations that require, for example, object manipulation or high precision, gesture interaction should be the main technique because voice commands can become inefficient and impractical if the required action is complex.

When using separate panels for buttons or other information, it is recommended that the user has the option to disable these panels from their view, or at least that they are movable and can be locked in place. We observed instances where our test users would have unintentional interactions with the menu buttons because the panel hovered too close to their hands. Adding many information panels can also easily clutter the AR space.

CONCLUSION

In this study, we explore the viability of two types of controls for HAR: voice commands and gestures, with the notion that in the near future, these may be the most viable interaction techniques for the persistent use of smart glasses when accessing the Metaverse.

Our findings show that gesture controls are found to be more natural and exciting, while voice commands are faster and more efficient. For this reason, we recommend that application designers and developers consider gesture interactions, especially for leisure and voice command interactions when efficiency and reliance are required. Both interaction techniques evoked some level of worry in users on how they would learn or remember how to use them but for different reasons. With gestures, these issues will be mitigated once the technology advances and the tracking techniques improve, however, with voice commands the users might still, in the future, require some type of memory assistance or tutorial implemented.

ACKNOWLEDGMENTS

This work was supported in part by European Commission under Grant NESTOR 101021851 and Grant IDUNN 101021911, in part by the Research Council of Finland

6G Flagship Program under Grant 346208, and in part by Enabling Metaverse under Grant 8719/31/2022.

The study followed the Ethics Committee of Human Sciences of University of Oulu, Finland and the Finnish Advisory Board on Research Integrity (TENK) guidelines on non-medical research projects, and Finnish and EU laws.^{26,27}

REFERENCES

1. T. Alsop, "AR headsets and glasses—Statistics & facts [Statistics]," Statista, 2023. [Online]. Available: <https://www.statista.com/topics/10134/ar-glasses>
2. R. Doerner et al., "Introduction to virtual and augmented reality," in *Virtual and Augmented Reality (VR/AR) Foundations and Methods of Extended Realities (XR)*. Cham, Switzerland: Springer, 2022, pp. 1–37.
3. R. Cheng, N. Wu, S. Chen, and B. Han, "Will metaverse be nextg internet? vision, hype, and reality," *IEEE Netw.*, vol. 36, no. 5, pp. 197–204, Sep./Oct. 2022.
4. S. Mystakidis, "Metaverse," *Encyclopedia*, vol. 2, no. 1, pp. 486–497, 2022.
5. S. S. Muhammad Nizam et al., "A review of multimodal interaction technique in augmented reality environment," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4-2, 2018, Art. no. 1460.
6. H. J. Kang, J.-H. Shin, and K. Ponto, "A comparative analysis of 3D user interaction: How to move virtual objects in mixed reality," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2020, pp. 275–284.
7. A. Khurshid et al., "Hand gesture recognition for user interaction in augmented reality (AR) experience," in *Virtual, Augmented and Mixed Reality: Design and Development: VAMR 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event*, 2022, pp. 306–316.
8. J. Toratti, "Integration of digital twin to augmented reality in industrial enrichment plants: An anticipated user experience study," Master's thesis, J. Toratti, University of Oulu, Oulu, Finland, 2022.
9. M. Prilla et al., "How to interact with augmented reality head-mounted devices in care work? A study comparing handheld touch (hands-on) and gesture (hands-free) interaction," *AIS Trans. Hum.-Comput. Interact.*, vol. 11, no. 3, pp. 157–178, 2019.
10. P. Bellani et al., "Enhancing user engagement in shared autonomous vehicles: An innovative gesture-based windshield interaction system," *Appl. Sci.*, vol. 13, no. 17, 2023, Art. no. 9901.

11. B. Pfleging et al., "Multimodal interaction in the car: Combining speech and gestures on the steering wheel," in *Proc. 4th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2012, pp. 155–162.
12. K. Pfeuffer et al., "PalmGazer: Unimanual eye-hand menus in augmented reality," 2023. [Online]. Available: <https://arxiv.org/abs/2306.12402>
13. M. Lee et al., "A usability study of multimodal input in an augmented reality environment," *Virtual Reality*, vol. 17, pp. 293–305, 2013.
14. S. Sadri et al., "Manipulating 3D anatomic models in augmented reality: Comparing a hands-free approach and a manual approach," in *Proc. IEEE Int. Symp. Mixed Augmented Reality*, 2019, pp. 93–102.
15. K. A. Satriadi, B. Ens, M. Cordeil, B. Jenny, T. Czauderna, and W. Willett, "Augmented reality map navigation with freehand gestures," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces*, 2019, pp. 593–603.
16. Microsoft, "HoloLens research mode," 2023. [Online]. Available: <https://learn.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/research-mode>
17. H. J. Guo and B. Prabhakaran, "HoloLens 2 technical evaluation as mixed reality guide," 2022. [Online]. Available: <https://arxiv.org/abs/2207.09554>
18. K. S. Hale et al., "Augmented reality sandtable (ARES) impacts on learning," *Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 63, no. 1, pp. 2149–2153, 2019.
19. A. Pukari et al., "LeARn flags: Assessment of the learning curve of inexperienced users with gesture interactions with hands-free augmented reality," in *Proc. 26th Int. Academic Mindtrek Conf.*, 2023, pp. 340–343.
20. Brooke, "SUS- A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
21. B. Laugwitz et al., "Construction and evaluation of a user experience questionnaire," in *HCI and Usability for Education and Work: USAB 2008*, Berlin Heidelberg: Springer, 2008, pp. 63–76.
22. V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006.
23. K. Malterud et al., "Sample size in qualitative interview studies: Guided by information power," *Qualitative Health Res.*, vol. 26, no. 13, pp. 1753–1760, 2016.
24. K. Lebeck, K. Ruth, T. Kohno, and F. Roesner, "Securing augmented reality output," in *Proc. IEEE Symp. Secur. Privacy*, 2017, pp. 320–337.
25. U. Riedlinger et al., "Tango vs. HoloLens: A comparison of collaborative indoor AR visualisations using hand-held and hands-free devices," *Multimodal Technol. Interact.*, vol. 3, no. 2, p. 23, 2019.
26. Finnish National Board on Research Integrity, "The ethical principles of research with human participants and ethical review in the human sciences in Finland," 2020. [Online]. Available: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf
27. University of Oulu, "Ethics committee of human sciences," 2024. [Online]. Available: <https://www.oulu.fi/en/university/faculties-and-units/eudaimonia-institute/ethics-committee-human-sciences>

MIKKO KORHAKOSKI is a researcher at the University of Oulu, 90570, Oulu, Finland. His research interests include situational awareness and multimodal interaction methods in augmented and virtual realities. Korhakovski received his M.Sc. degree in computer science and engineering from the University of Oulu, Finland. He is a member of IEEE and ACM. He is the corresponding author of this article. Contact him at mikko.korkiakoski@oulu.fi.

PAULA ALAVESA is a researcher and a university lecturer with the University of Oulu, 90570, Oulu, Finland. Her research interests include technologies, games, and other applications that combine digital, virtual and physical spaces. Alavesa received her Ph.D. degree in computer science from the University of Oulu, Finland. She is a member of IEEE, ACM, SIGCHI, and DiGRA. Contact her at paula.alavesa@oulu.fi.

PANOS KOSTAKOS is a senior research fellow with the University of Oulu, 90570, Oulu, Finland. His research is situated at the intersection of AI, Information Security and Security orchestration, focusing on autonomous, mutable, and cognitive cyber defense mechanisms. Kostakos received his Ph.D. degree in criminology from the University of Bath, Bath, U.K. Contact him at panos.kostakos@oulu.fi.