VUM: Understanding Requirements for a Virtual Ubiquitous Microscope

Francisco Kiss
University of Stuttgart
Stuttgart, Germany
franciscokiss@gmail.com

Paweł W. Woźniak
Utrecht University
Utrecht, the Netherlands
p.w.wozniak@uu.nl

Verena Biener
Coburg University of Applied Sciences and Arts
Coburg, Germany
verena.biener@hs-coburg.de

Pascal Knierim
LMU Munich
Munich, Germany
pascal.knierim@um.ifi.lmu.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@um.ifi.lmu.de

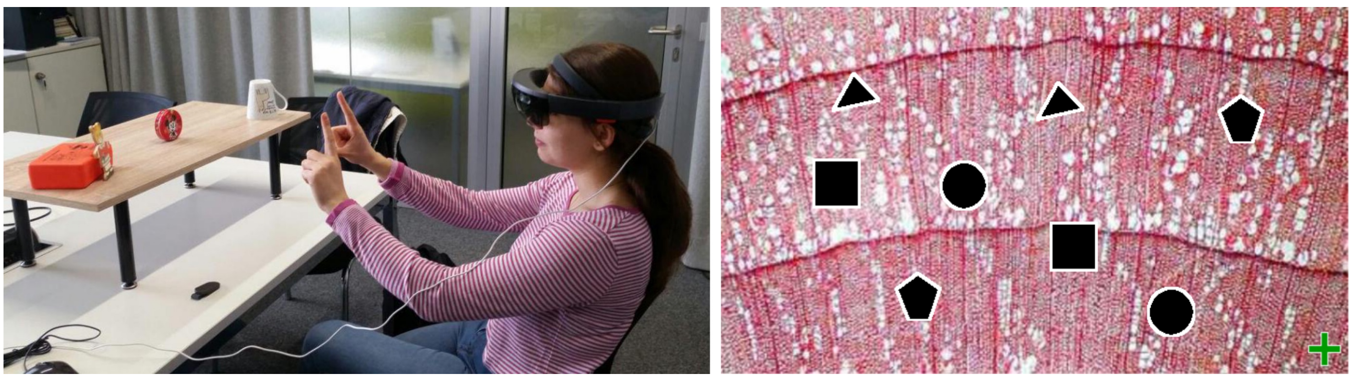


Figure 1: Prototype of a Virtual Ubiquitous Microscope to investigate interaction techniques.

ABSTRACT

Augmented Reality (AR) and wearable sensors offer new possibilities to expand our senses and change how we interact with the world. Sensory augmentation can be integrated into everyday activities, but controls remain a challenge for user experience. In this paper, we investigate how users can control a futuristic interface that enables in-situ magnification. We designed an interactive system to enable users to zoom in on objects up to a microscopic level and implemented a prototype using the Microsoft Hololens. In a user-study, we compared full-screen to windowed visualizations and four interaction techniques for zooming: a clicker, two types of gestures, and voice. Our results indicate that the clicker enabled users to zoom at the fastest rate and lowered cognitive load. We also found a preference for windowed views. With our work, we provide insights for future augmented vision systems.

MUM 2020, November 22–25, 2020, Essen, Germany

ACM ISBN 978-1-4503-8870-2/20/11…$15.00
https://doi.org/10.1145/3428361.3428386

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; **Interactive systems and tools**; **Gestural input**.

KEYWORDS

virtual microscope, augmented reality

ACM Reference Format:

Francisco Kiss, Paweł W. Woźniak, Verena Biener, Pascal Knierim, and Albrecht Schmidt. 2020. VUM: Understanding Requirements for a Virtual Ubiquitous Microscope. In *19th International Conference on Mobile and Ubiquitous Multimedia (MUM 2020), November 22–25, 2020, Essen, Germany*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3428361.3428386

1 INTRODUCTION

Humans have always used technology to enhance their sensory perception. Starting with simple monocles, today's modern technologies enable extraordinary abilities to enhance human vision, such as seeing beyond the naturally visible spectrum of light [1], or decelerating the speed of what we visually perceive [12]. We expect this trend to continue, offering ever-increasing levels of augmentation and empowering users to perceive much more than typically possible.

The magnification of vision is an established human need. Historical sources report on ancient attempts of overcoming the natural limits of vision through the use of lenses [26]. Since the 17th century, the use of microscopes, and for a few decades electronic microscopes, has allowed humans to perceive objects several orders of magnitude smaller than what the naked eye can see. Scientists, and to some limited extent the general population, have used these tools to great benefit. Ubiquitous access to different levels of visual magnification can be useful in many situations, both for technical and more trivial activities. Humans interact with small objects on a daily basis. Being able to take a closer look at small details can have a positive effect on a broad range of activities.

With the advent of ubiquitous and wearable computing, the incorporation of microscopy to Augmented Reality (AR) becomes foreseeable, or at least feasible [27]. See-through Head-Mounted Displays (HMD) offer the novel opportunity of presenting visual information ubiquitously, dynamically, and on-demand. They can enable users to magnify objects in their immediate surroundings at will. This kind of augmentation is aligned with Schmidt's vision of seamless integration of sensory amplification and enhancement into users' lives, with users perceiving the amplification not as a tool, but as a direct augmentation of their perception [24]. This kind of interaction, ubiquitous and seamless, poses a series of challenges quite different from those posed by traditional *ubicomp* interfaces. Screens with icons, text, or diagrams become an obstacle when the user only wishes to see better. Traditional interaction metaphors become rapidly cumbersome when they interrupt what we usually do intuitively in a highly automated way. Augmenting the human senses should make perception *easier* and *better*, so controlling augmented vision should not be more complicated than controlling natural vision, nor require more effort. From an HCI perspective, the study of a Virtual Ubiquitous Microscope (VUM) offers an exiting probe into sensory augmentation and an opportunity to gain a better understanding of how to design interfaces for this emerging field.

In this paper, we conduct an exploratory probe in the design of interfaces for sensory augmentation using a VUM as a functional example. Recognizing the difficulty of an abrupt shift in interaction paradigms, we take an incremental approach, assessing which existing methods are perceived more favorably by users. For this purpose, we designed an AR system for the ubiquitous visualization of magnified imagery and conducted a user study comparing different input and output techniques. Admittedly, the development of a VUM presents an additional challenge that will not be addressed in this work, namely the technical implementations of sensors that enable real-time magnification. Our contribution is an interaction design for a VUM and the evaluation of four different interaction techniques and two different display techniques for such design. In the next section, we review past work and identify the critical aspects of the design for a VUM. Next, we present our design and describe the proposed interaction and display techniques. We follow by reporting on the experiment design and discuss its results. This paper concludes with a description of the limitations of this work and a conclusion.

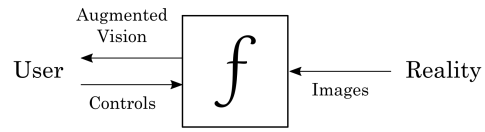


Figure 2: Functional model of the interaction with a Virtual Ubiquitous Microscope: the system receives input from the user in form of controls and from reality in form of images. The output of the system is the magnified image.

2 RELATED WORK

Sensory augmentation can be modeled as a mediation of human perception [16, 22]. In the case of visual amplification, such mediation would consist of a filter affecting the amount of detail and field-of-view a user can see for a given object. For the particular case of a VUM, it is thus possible to model the interactive system as a function that receives as input the images to be magnified and the user controls, and returns the magnified image (see Figure 2).

To our knowledge, there is no previous work in the literature that investigates the challenges of interaction design for virtual microscopes. However, the visualization of magnified images has been subject of HCI research, and previous work in this area provides a basis for our work, both for the input and output characteristics of the system.

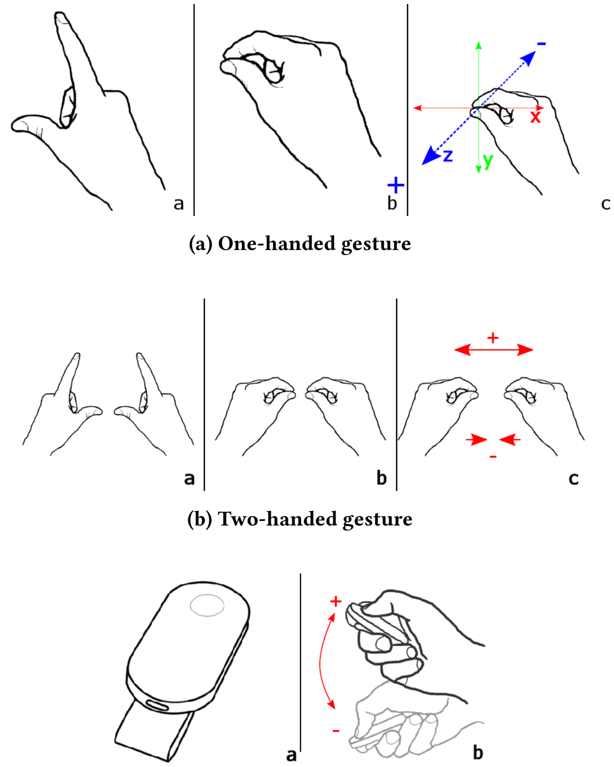
The dynamic display of magnified images has multiple aspects that require attention. The transition between normal to magnified vision, as well as between levels of magnification, can be performed in a continuous manner, or by incremental discrete intervals. Work by Bartram et al. suggests that continuous transitions benefits the understanding of information arranged hierarchically [3] while research by Chen et al. [7] and Pan et al. [20] provide support for its technical feasibility and highlight its relevance.

The usage of optical-see-through HMD offers a versatile platform for the augmentation of vision but prompts a series of design decisions. It is necessary to opt between displaying the images using the complete field-of-view supported by the device or to present the user with a delimited area displaying the image. Further, the images must be positioned respect a coordinate system that can either be the real world, or the centered in the user. This has significant implications for the behavior of the interaction when the user moves. There is no specific research for the case of virtual microscopes, but it is possible to gain some insights from past research on map representation and navigation. Previous work focuses on both the use of floating frames (lenses) to display delimited regions of magnified images [15], and use of full field-of-view for immersive zooming interfaces [19]. Both approaches present benefits and disadvantages and thus it is unclear which would result in a better user experience for the particular case of a VUM.

Another aspect of the interaction is *control*. Controlling magnified views presents singular challenges, both due to the difference between granularity of movements and different levels of magnification [2]. Past work proposes the control of zooming interfaces with tangible controllers [5, 17], and one-handed or two-handed gestures [10, 21].

Satriadi et al. explored the interaction space for AR and VR using gestures and handheld controllers [23]. Through user studies, the authors compared different control possibilities for maps in different formats, including flat maps, curved ones, and globes. The central contribution of this work resides in two techniques to investigate hybrid input for mid-air gestures. Although this work was conducted around a clearly different application, namely map navigation, the general approach used by the authors presents many inspiring elements. Dünser and Billinghurst recognized a number of challenges in the evaluation of AR systems and applications, and examined common evaluation techniques used in user studies [8]. This paper uses the methodology proposed by Dünser and Billinghurst.

Past work highlights the necessity for a better understanding of input and output methods for sensory augmentation. Thus, our contribution to this area of research consists on a research probe aligned with the goals and methods of the literature and applied to the concrete case of a virtual ubiquitous microscope. Further, we provide general insights for interaction designs in line with the emerging new paradigm of sensory augmentation.



(a) One-handed gesture

(b) Two-handed gesture

(c) Physical controller

Figure 3: Four Interaction Techniques were evaluated to control the VUM: (a) one-handed gesture, (b) two-handed gesture, (c) a physical controller, and voice commands (not depicted).

3 VIRTUAL UBIQUITOUS MICROSCOPE

To assess the interaction challenges of ubiquitous augmentation, we designed, implemented, and evaluated a prototype of a VUM. We emphasize that the main contribution of our work is not the design itself, but the evaluation of different interaction methods, with the VUM serving only as a vehicle for our study.

A VUM has a defined functionality: when activated, it shows the user a magnified view of a particular object of interest. Further, the user should be able to control the level of magnification, or terminate the interaction and return to normal vision.

To simplify the interaction we defined two design constraints: First, the interaction must be started only intentionally by the user and thus avoiding accidental activation is necessary. Second, the operation of the system must in no way hinder the performance of the task in which the system aids the user. Consequently, the activation and termination, as well as the control of the level of magnification, need to be performed with simple, low effort actions.

Traditional microscopes do not offer the possibility of panning, even if such functionality would be desirable [2]. Given the level of magnification of a microscope, navigation is difficult to control, since even very small displacements result in the object of observation disappearing from the field of view of the device. Thus, the VUM avoids the problem of panning by presenting the magnification of the static image recorded at the beginning of the interaction.

These design choices allowed us to focus on the principal goals of our investigation: the Interaction and Display Techniques for VUM. Based on past work, we proposed four Interaction Techniques and two Display Techniques, which are exemplary of typical approaches in AR and VR applications. This comparison between different proposed methods was recommended by Dünser and Billinghurst for novel AR applications, given the lack of well-established frames of reference for such evaluations [8].

3.1 Interaction Techniques

We selected four Interaction Techniques which represent current approaches in interacting with AR: one- and two-handed gestures, a tangible input device, and voice commands. The first three are derived from previous work [5, 10, 21]. We also included voice commands since the efficiency over gestures for some particular cases led us to consider this technique a likely Interaction Technique in future AR interaction designs [13].

We chose the specific gesture for each Interaction Technique based on the approaches used in the discussed related work while prioritizing keeping the operation of the device efficient.

We defined the interaction with a VUM as having three stages: first, the VUM is activated, and the interaction starts. The next stage is viewing of the magnified object while controlling the zoom level. Finally, the interaction can be terminated by zooming out beyond the minimal level of magnification.

3.1.1 One-Handed Mid-Air Gestures. The one-handed mid-air gesture aims to resemble the movement a person would perform to move an object closer or further away from the eyes, to examine it in more detail or gain a more general overview. The user starts the interaction by raising the index finger and pressing down the finger towards the thumb. While keeping the fingers together, the user can move the hand along the z-axis, which means closer or further away from the body, to change the magnification level. This Interaction Technique is illustrated in Figure 3a. This particular hand gesture was selected due to the intuitive and controllable relationship between the performed distance and the level of zoom. Additionally, this gesture is commonly used in AR platforms, such as the Microsoft Hololens[1].

3.1.2 Two-Handed Mid-Air Gestures. Two-handed mid-air gestures offer a more extensive design space for interaction. However, performing two-handed gestures adds extra effort and the impossibility of performing a parallel task with the other hand. Chaconas and Höllerer [6] compared different two-handed gesture sets for rotation and scaling and comparing them to a one-handed gesture set. Based on their findings, we choose the two-handed gesture to manipulate the magnification level of the AR microscope. Similarly to the one-handed gesture, users must raise both index fingers and press subsequently against the thumb to start the interaction. While holding the fingers together, the user can move the hands apart or together in order to modify the magnification level of the virtual microscope. This Interaction Technique is illustrated in Figure 3b. This gesture was inspired by previous work [10] and, similarly to the one-handed gesture, due to the relationship between the separation of the hands and the level of magnification. Further, it mimics a commonplace conversational gesture used to emphasize size.

3.1.3 Physical Controller. A physical controller represents the ability of the user to control a system and gives the user direct haptic feedback for command inputs. Further, this Interaction Technique is more flexible respect its posture requirements and users can operate the system without keeping their arms mid-air, thus reducing physical fatigue. To change the magnification level, users press a button and rotate the tip of the controller up or down. The controller movement is illustrated in Figure 3c. For this choice of gesture we consciously avoided using a traditional mouse to remain consistent with the other mid-air gestures. The particular choice of gesture considers typically tangible controllers, which commonly feature accelerometers and gyroscopes. Such sensors can readily measure rotational movements yet cannot quantify linear distances with precision.

3.1.4 Voice Command. As the last Interaction Technique, we enable users to operate our prototype through voice commands. The user can command the system with speech to change the magnification level. By saying "smaller", users can lower the magnification level, while "bigger" increases the magnification. The exact increment and decrement of magnification levels depend on the implementation and are discussed in the Apparatus section.

3.2 Display Techniques

We designed two different Display Techniques to show the magnified content to the user in line with previous work [15, 19]. On one hand, we developed the *head-lock* visualization. For this technique, the magnified content covers the entire available display space. On the other hand, the *tag-along* visualization displays the magnified content in a rectangular window floating in mid-space and positioned on top of the observed object. Although further alternative techniques would be possible, this selection reflects the contrasting approaches of augmenting the *observer* and augmenting the *observed*. This dichotomy is emphasized by the first technique completely immersing the user in the interaction, while the second technique limits the interaction interface in size and position. Further, these two Display Techniques are consistent with our exclusion of panning and navigation. Given the difficulty in finding a compromise between the two discussed Display Techniques, their comparison can yield useful insights for future design. In the following paragraphs, we explain each technique in detail.

3.2.1 Head-Lock Visualization. With the *head-lock* technique, the entire available display-space offered by the AR HMD is used to visualize the magnified content, frequently called *full-screen*. When the VUM is activated, the complete field of view of the user is filled by the magnified image. This visualization is locked to the user's head and follows the user's movements anywhere. Hence, we avoid forcing the user to remain completely still during the observations. The user can navigate through the different levels of magnification with the designated Interaction Technique and move his head around freely without altering the focus of magnification (see Figure 4a). The advantage of this technique is the large display area, enabling a wider oversight of the magnified image. However, this mechanism hinders the user from diverging attention from the magnified image and impedes moving around safely. This translates to the inability to perform other tasks while zooming into things.

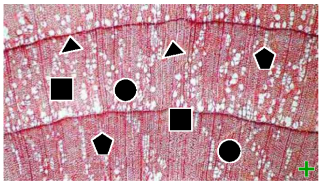
3.2.2 Tag-Along Visualization. In contrast to *head-lock*, the *tag-along* technique for visualization has a small virtual display area that is attached to the position of the observed object. When the user activates the ubiquitous microscope and zooms into an object, a postcard-sized virtual display appears in front of that object, showing its microscopic magnification (see Figure 4b). The virtual display remains stationary in space and does not follow the users' movements. Therefore, the user can look in another direction, and the magnified image will stay where it was and save its current status. This allows the user to magnify multiple objects in parallel and switch between them, regaining control of a magnified view by focusing the gaze on it. In consequence, each instance of the magnification view should be terminated individually.

The main advantage of this display technique is the freedom of the user to control and compare multiple magnification views of objects, while also being able to interact with the real world. This comes at the price of a less immersive experience and a limited display area, making the observation of smaller details or the comparison of features within a particular level of magnification less effective.

4 EVALUATION

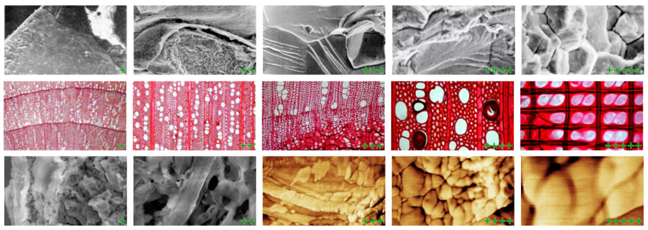
We evaluated the performance and user preference of our design using a functional prototype. Further, we compared the different Interaction and Display Techniques in terms of these two aspects. In the following, we describe our methodology and experimental design, and present the collected data.

[1]https://docs.microsoft.com/en-us/windows/mixed-reality/gestures



(a) Task: count the squares across all levels of magnification (here for the *head-lock* Display Technique).

(b) Floating frames of the *Tag-along* Display Technique.



(c) Images used for all levels of magnification for metal (top row), wood (middle row), and ceramic (bottom row). Image sources: [9, 18, 25]

Figure 4: Display Techniques (top) and image set used for the user study (bottom).

4.1 Methodology

We used a 4*x*2 within-subject experiment design with the independent variables Interaction Technique and Display Technique. For Interaction Techniques, we had four levels: *one-handed*, *two-handed*, *controller*, and *voice*. For Display Techniques we had two levels, *head-lock* and *tag-along*. The performance was measured while participants interacted with the prototype to solve a visual search task across different levels of magnification. We recorded the task completion time (TCT), training time, and error rate, as well as task workload [11], and personal preference once each task was finished.

4.2 Apparatus

For this experimental study, eight different versions of the prototype were implemented; one for each combination of Interaction and Display Techniques. These were implemented in Unity[2] and presented to users on a Microsoft HoloLens[3]. The physical controller Interaction Technique was performed using the HoloLens clicker.

To simulate the magnification functionality, we implemented an object recognition algorithm using Vuforia[4]. This was trained for a small set of objects of different materials: a wooden toy, a metal candy box, and a ceramic cup. These objects were chosen for being distinctive yet familiar items that are typically made of the material they represent. The system can individually recognize these objects in the center of the user's field of view. To aid the user in the selection process, a green dot was displayed at the center of the field of view to indicate that the object was recognized and the magnification could be initiated for that object.

Due to the inherently discrete nature of the voice Interaction Technique, allowing only step-wise increments and decrements, we opted to implement the magnification using fixed discrete levels. This also simplifies the prototype and ensures consistency across trials and conditions, since all users will experience the same levels of magnification, and thus reduce the distortions caused by *quid tertium*. We simulated the different levels of magnification using a set of static images [9, 18, 25] with different levels for each material (see Figure 4c).

The discrete level of magnification was matched to the continuous distance of the movement performed by the user either with the hands or using the physical controller. The ratio between moved distance and level of magnification was determined empirically through a pilot study conducted specifically for this purpose. This way we determined comfortable ranges of movement for the three physical Interaction Techniques. For the *One-handed* gesture, the minimal level of zoom was position where the interaction starts, which is the distance of a semi-extended arm, and the maximal level of zoom was reached for the hand positioned around 10 centimeters away from the face of the user. For the *Two-handed* gesture, the minimal level of zoom was fixed to the initial distance between the user hands, 20 centimeters, and the maximal level of zoom was reached around shoulder-width, 50 centimeters. For the rotation of the *Controller*, the minimal level was fixed to the starting position and the maximal for a rotation of 90 degrees respect the initial angle. For all three methods, the intermediate levels of magnification with distributed at regular intervals between the minimum and maximum.

Using the corresponding interaction technique for each condition, the images were displayed when the interaction was initiated, starting with the lowest level of magnification. The user would control the level of magnification by switching between images using the different Interaction Techniques. To give the users a reference for the current level of magnification, it was visually represented with increasing + symbols in the bottom right corner of the image. Finally, when the user terminated the interaction by zooming out beyond the minimum level of magnification, the magnified image would disappear.

4.3 Task

To investigate the effects of the Interaction and Display Techniques, participants were asked to perform a task that required them to navigate through all levels of zoom and retrieve information from the magnified images. For this purpose, we designed a visual search task based on previous work, requiring the identification of shapes [4]. Each level of magnification for each material (5 levels per object, a total of 15 images) was marked with a varying amount of black geometric figures (squares, circles, triangles, and pentagons). Each image contained a total of eight symbols, with the quantity of each shape being different for each image (see Figure 4a). The objective of the task was to count the total number of squares present across the different levels of magnification for each material. For each condition, one of eight sets of marked images was assigned in a counterbalanced fashion, with each set having

[2]https://store.unity.com/de/products/unity-personal
[3]https://www.microsoft.com/en-us/hololens
[4]https://library.vuforia.com/articles/Training/Object-Recognition

a total number of squares ranging from 13 to 15. To minimize the chances of participants guessing the number, the range of possible solutions was unknown to the participants. With this design, each participant had to check the shape of a total of 120 objects and keep track of the squares, while controlling the level of magnification and switching objects.

4.4 Procedure

Each participant was first briefed and asked for consent and demographic information. During the experiment, the participant remained sitting on a chair. The three objects were placed in front of the participant on a small table. The participant was asked to wear the HoloLens and adjusted it to fit comfortably. The trial task was then explained to the participant. A prototype version with unmarked images was loaded and initiated on the HoloLens using a laptop. This initial set of images did not present geometric shapes and was intended to help the participant get familiarized with the HoloLens and the Interaction and Display Techniques for the current condition. The participant was encouraged to test the prototype until they felt comfortable interacting with it. The time needed for training was recorded and then the application for the actual task was started, this time with images presenting the geometric shapes. At this point, the stopwatch was restarted. The participant then proceeded to count the squares and upon finishing, reported their total number. This result and the TCT were noted and the participant was given a NASA TLX questionnaire to fill. These steps were repeated until all four interaction techniques for a given visualization were done. Next, this procedure was repeated for the other display technique. The order of the Interaction Techniques and Display Techniques were individually counterbalanced across participants, with the constraint of clustering conditions by display technique to avoid confusing the participant.

At the end of each trial, a short semi-structured interview was conducted, were participants indicated their preference for an interaction technique and explained the reasons for this preference. Additionally, participants were encouraged to suggest application scenarios for a VUM. The total participation time ranged from 30 to 60 minutes.

4.5 Participants

We recruited 16 participants. Five of them identified themselves as females and eleven as males, and their age ranged between 18 and 27 years. All participants reported being in perfect health, presenting normal or corrected vision (with glasses or contact lenses). Eleven participants had no prior experience with the Microsoft HoloLens, four had little experience, attained through other studies in which they participated. One participant stated to be an experienced user.

4.6 Quantitative Results

A two-way ANOVA was conducted to examine the effect of Interaction and Display Techniques on TCT. The test showed a significant effect of Display Technique $F(3, 119) = 6.30$, $p < .001$ and Interaction Technique $F(1, 119) = 12.41$, $p < 0.001$. There was no significant interaction effect, $F(3, 119) = 1.61$, $p = .19$. Post-hoc Tukey HSD for the TCT showed significant differences between *one-handed* and *controller* ($p < .01$), and *controller* and *two-handed* ($p < .01$). Post-hoc comparisons of individual Display Technique × Interaction Technique pairs are shown in Figure 5a. Group mean and standard deviation are shown in Table 1.

Table 1: Mean value, standard deviation, and standard error for Task Completion Time and Error, calculated for each condition of the two variables.

	TCT		Error	
	Mean	SD	Mean	SD
One-handed	121.31	79.63	0.38	0.87
Controller	71.97	21.47	0.31	0.59
Two-handed	124.06	84.91	0.56	0.95
Voice	95.69	27.22	0.31	0.54
Head-lock	120.42	79.88	0.50	0.87
Tag-along	86.09	34.53	0.28	0.60

We investigated the effect of the two factors on error rate using a two-way ANOVA. We observed no significant effects to be caused by either Display Technique, $F(3, 119) = 0.80$, $p = .45$, or Interaction Technique, $F(1, 119) = 2.75$, $p = .10$.

A two-way ANOVA which investigated the effect of Display Technique and Interaction Technique on the NASA TLX score revealed a significant effect of Interaction Technique: $F(1, 119) = 24.47$, $p < .001$. There was no effect of Display Technique, $F(3, 119) = 0.75$, $p > .05$ nor an interaction effect, $F(3, 119) = 0.76$, $p > .05$. Post-hoc Tukey HSD results are shown in Figure 5c.

4.7 Qualitative Results

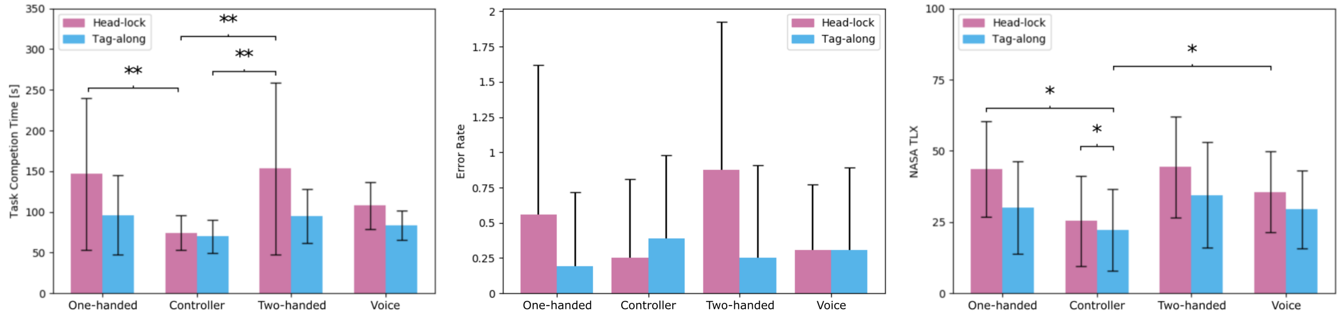
We interviewed participants about their preference for Interaction and Display Techniques, and prompted them to envision applications scenarios. The preferred Interaction Technique was the *controller*, being the first choice for 12 from 16 of participants. Both *one-handed* and *voice* techniques were chosen by two participants, and *two-handed* was chosen by no participant. When asked about the least preferred method, *one-handed* and *two-handed* gestures were chosen by six participants each, and *voice* by four. The preference for the *controller* was mainly attributed to ease-of-use by nine of interviewees, to reliance by nine, and to intuitive control of the interaction by six.

The *tag-along* Display Technique was preferred by nine participants. Participants explained this preference by noting the ability to see multiple images simultaneously. On the other hand, interviewees who preferred *head-lock* explained this to be caused by the image following the gaze of the user, allowing freedom of movement, and that improved the viewing in terms of degree of detail and contrast.

The application scenarios envisioned by the interviewees were in the areas of clinical medicine, technical design, research in biology, materials science, and education.

5 DISCUSSION

The collected results suggest that user appreciated both Display Techniques, but *tag-along* performed significantly better in terms of TCT. The Interaction Techniques also had a significant effect



(a) **(b)** **(c)**

Figure 5: Mean value and standard error of task completion time (a), error rate (b), and task load index score (c) for the four Interaction Techniques and two Display Techniques. Post-hoc significance is indicated (* p<0.05, ** p<0.01)

on the TCT, and show that using the *controller* is more effective other techniques. The rate of errors committed in the task provides little insight, only indicating better performance of *tag-along* visualization technique. The difference in task load was significant across Interaction Techniques, reinforcing the preference for the *controller*.

Based on these observations, we conclude that physical controllers are the preferred choice for controlling a VUM. This technique results in better performance and lower task load. While the study produced a clear result, we do not advise that future systems resign from developing alternative controls. Part of the results of the experiment can be explained by the fact that the remote may be the device which felt most familiar to the majority of the users. Perhaps, this is a case of legacy bias in both performance and preference. This might also explain, to some extent, the preference for the *head-lock* mode by some users despite better performance when using the *tag-along* display. Enabling users to switch between modes can bring further understanding on this topic, since it remains unclear if the effects in performance and preference are due to human physiological and perceptual characteristics. Arguably, design shapes perception as much as perception shapes design. Users trained and formed within a paradigm may develop preference and expertise within that paradigm. This suggests the need to introduce incremental steps in the transition to new interaction paradigms, such as augmented perception. Users performed better with physical controllers and voice commands, which is aligned with the common practices for controlling devices (e.g. TV remote) or asking people to perform an action.

Quantitative results indicate better performance for the *tag-along* technique. However, a large part of the participants preferred the *head-lock* technique. The results suggest that the visualization mode has a stronger impact on the interaction for the given approaches, while users easily adapt to different Interaction Techniques, obtaining similar performance and proficiency in controlling the system. This observation is supported by the error rate, TCT, and NASA TLX scores. These results are aligned with the findings by Lee et al. [14], confirming that the variation of typical Interaction Techniques has a limited effect in performance during the interaction. However, we must recognize that the selection of investigated Interaction and Display Techniques is by no means exhaustive, and thus the generalization of our findings needs to be applied with caution. On the other hand, this kind of limitation is typical for the evaluation of physical prototypes of AR applications, as discussed by Dünser and Billinghurst [8].

From an interaction design perspective, our findings highlight the challenges for interfaces in the age of ubiquitous computing. Design decisions such as visualization or control methods have a profound impact on the quality of the interaction, thus signaling the importance of *good design* in the successful incorporation of AR to human activities as suggested by Schmidt [24].

Consequently, future work on VUM controls should investigate combining both Display Techniques, enabling users to toggle between the views, and investigate the users' choices in different use cases. Many challenges remain to be addressed, such as the incorporation of navigation and panning in the interaction. Alternative display methods are still unexplored. Semi-transparent displays or enabling a transition between the two modes studied in this work offer further challenges to HCI.

6 LIMITATIONS

This research contributed an exploratory AR application, which particularly focused on sensory enhancement and augmentation. Thus, we investigated a limited subspace of the whole interaction design space for AR, namely only four input and two output methods. Further, the particular implementation of the methods is a research prototype, limited in terms of versatility and flexibility, such as forcing the users to start two-handed gestures with their hands apart. We designed the application as an isolated concept. If VUM were to be part of a generic AR interface, the selected techniques might be already assigned to control other functions. Finally, we cannot exclude the possibility of the discrete magnification levels affecting the interaction and this aspect needs to be addressed in future work.

7 CONCLUSION

In this paper, we explored interaction and display techniques for a VUM. We proposed possible techniques that could help incorporate this system into human perception, following the vision of Sensory

Amplification. We designed the interaction in two dimensions: interaction and display. We proposed four Interaction Techniques, one- and two-handed gestures, a physical controller, and voice commands, and two Display Techniques, one fixed to the user in a full-screen mode, and one fixed to the magnified object inside a floating frame. We implemented a functional prototype based on the Microsoft HoloLens that incorporates all four interaction modalities. We conducted a user study with 16 participants. In a search task, participants had to recognize and count squares among other shapes through different levels of magnification. Our results indicate a increased performance for *tag-along* visualizations and input based on tangible controllers. This advantage is also reflected by user preference in both cases. However, both Display Techniques was popular among users despite the differences in performance, suggesting that both modes are equally desirable. We hope that our work will inspire further research into how current developments in AR can be used to enable sensory augmentation.

ACKNOWLEDGMENTS

This research was supported by the European Unions Horizon 2020 Programme under ERCEA grant no. 683008 AMPLIFY.

REFERENCES

[1] Yomna Abdelrahman, Pawel Wozniak, Pascal Knierim, Niels Henze, and Albrecht Schmidt. 2018. Exploration of Alternative Vision Modes Using Depth and Thermal Cameras. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (Cairo, Egypt) *(MUM 2018)*. ACM, New York, NY, USA, 245–252. https://doi.org/10.1145/3282894.3282920

[2] Caroline Appert, Olivier Chapuis, and Emmanuel Pietriga. 2010. High-precision Magnification Lenses. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. ACM, New York, NY, USA, 273–282. https://doi.org/10.1145/1753326.1753366

[3] Lyn Bartram, Albert Ho, John Dill, and Frank Henigman. 1995. The continuous zoom: A constrained fisheye technique for viewing and navigating large information spaces. In *Proceedings of the 8th annual ACM symposium on User interface and software technology*. 207–215.

[4] Stefanie I Becker, Anthony M Harris, Dustin Venini, and James D Retell. 2014. Visual search for color and shape: When is the gaze guided by feature relationships, when by feature values? *Journal of Experimental Psychology: Human Perception and Performance* 40, 1 (2014), 264.

[5] Frédréric Bourgeois and Yves Guiard. 2002. Multiscale Pointing: Facilitating Pan-zoom Coordination. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. ACM, New York, NY, USA, 758–759. https://doi.org/10.1145/506443.506583

[6] N. Chaconas and T. Höllerer. 2018. An Evaluation of Bimanual Gestures on the Microsoft HoloLens. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1–8. https://doi.org/10.1109/VR.2018.8446320

[7] Hsueh-I Chen, Claire Chen, and Yung-Yu Chuang. 2017. A Deep Convolutional Neural Network for Continuous Zoom with Dual Cameras. In *SIGGRAPH Asia 2017 Posters* (Bangkok, Thailand) *(SA '17)*. ACM, New York, NY, USA, Article 13, 2 pages. https://doi.org/10.1145/3145690.3145717

[8] Andreas Dünser and Mark Billinghurst. 2011. Evaluating augmented reality systems. In *Handbook of augmented reality*. Springer, 289–307.

[9] Mohammadamin Emami, Yilmaz Sakali, Chiristian Pritzel, and Reinhard Trettin. 2016. Deep inside the ceramic texture: A microscopic chemical approach to the phase transition via partial-sintering processes in ancient ceramic matrices. *Journal of Microscopy and Ultrastructure* 4, 1 (2016), 11 – 19. https://doi.org/10.1016/j.jmau.2015.08.003

[10] Vito Gentile, Daniele Fundarò, and Salvatore Sorce. 2019. Elicitation and Evaluation of Zoom Gestures for Touchless Interaction with Desktop Displays. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (Palermo, Italy) *(PerDis '19)*. ACM, New York, NY, USA, Article 16, 7 pages. https://doi.org/10.1145/3321335.3324934

[11] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[12] Pascal Knierim, Thomas Kosch, Gabrielle LaBorwit, and Albrecht Schmidt. 2020. Altering the Speed of Reality? Exploring Visual Slow-Motion to Amplify Human Perception using Augmented Reality. In *Proceedings of the 1st Augmented Humans International Conference (AHs '20)*. ACM, New York, NY, USA. https://doi.org/10.1145/3384657.3384659

[13] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. 2013. A Usability Study of Multimodal Input in an Augmented Reality Environment. *Virtual Real.* 17, 4 (Nov. 2013), 293–305. https://doi.org/10.1007/s10055-013-0230-0

[14] Minkyung Lee, Mark Billinghurst, Woonhyuk Baek, Richard Green, and Woontack Woo. 2013. A usability study of multimodal input in an augmented reality environment. *Virtual Reality* 17, 4 (2013), 293–305.

[15] Julian Looser, Mark Billinghurst, and Andy Cockburn. 2004. Through the Looking Glass: The Use of Lenses As an Interface Tool for Augmented Reality Interfaces. In *Proceedings of the 2Nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia* (Singapore) *(GRAPHITE '04)*. ACM, New York, NY, USA, 204–211. https://doi.org/10.1145/988834.988870

[16] Steve Mann. 2002. Mediated reality with implementations for everyday life. *Presence Connect* 1 (2002).

[17] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. 2019. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 585, 11 pages. https://doi.org/10.1145/3290605.3300815

[18] Martin Möser. 1982. 4. Elektronenmikroskopische Fraktographie. (01 1982). https://doi.org/10.1007/978-3-642-93211-3_14

[19] Alessandro Mulloni, Andreas Dünser, and Dieter Schmalstieg. 2010. Zooming Interfaces for Augmented Reality Browsers. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) *(MobileHCI '10)*. ACM, New York, NY, USA, 161–170. https://doi.org/10.1145/1851600.1851629

[20] Kuang-Yu Pan and Yung-Yu Chuang. 2016. Continuous Zoom with Two Fixed-focal-length Lens. In *SIGGRAPH ASIA 2016 Posters* (Macau) *(SA '16)*. ACM, New York, NY, USA, Article 14, 2 pages. https://doi.org/10.1145/3005274.3005309

[21] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. Partially-indirect Bimanual Input with Gaze, Pen, and Touch for Pan, Zoom, and Ink Interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 2845–2856. https://doi.org/10.1145/2858036.2858201

[22] Jun Rekimoto and Katashi Nagao. 1999. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of User Interface Software and Technology'95*.

[23] Kadek Ananta Satriadi, Barrett Ens, Maxime Cordeil, Bernhard Jenny, Tobias Czauderna, and Wesley Willett. 2019. Augmented reality map navigation with freehand gestures. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 593–603.

[24] Albrecht Schmidt. 2017. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing* 16, 1 (2017), 6–10.

[25] Werner Schoch, Iris Heller, Fritz Schweingruber, and Felix Kienast. 2004. Wood anatomy of central European Species. (2004). www.woodanatomy.ch

[26] George Sines and Yannis A Sakellarakis. 1987. Lenses in antiquity. *American Journal of Archaeology* (1987), 191–196.

[27] Eric Tremblay, Igor Stamenov, Ralph Beer, Ashkan Arianpour, and Joseph Ford. 2013. Switchable telescopic contact lens. *Optics express* 21 (07 2013), 15980–15986. https://doi.org/10.1364/OE.21.015980