Extended Reality User Interface Quality of Experience Testing for Military Purposes Using a Video Game

Katarina Mišura, Miro Čolić* and Mirko Sužnjević
University of Zagreb Faculty of Electrical Engineering and Computing
Department of Telecommunications
Multimedia Quality of Experience Research Laboratory
CROATIA

katarina.misura@fer.unizg.hr; miro.colic@fer.hr mirko.suznjevic@fer.unizg.hr

* Croatian Defence Academy "Dr. Franjo Tuđman", Zagreb,
CROATIA

ABSTRACT

We developed a video game simulation of an augmented reality (AR) Heads-Up Display (HUD) system to evaluate technology acceptance and Quality of Experience (QoE) without the constrictions of current XR headsets. The simulation environment allows us to effectively iterate on conceptual designs and ideas according to qualitative feedback from participants. Given the complexities soldiers face navigating urban environments with multiple threats, we explore the potential of an AR HUD system to enhance navigation and information reception in urban warfare settings. In the simulation, we equipped soldiers with an XR HUD that provides navigational aids and additional data.

To assess the concept, participants performed a route-following task in a simulated urban environment using the AR HUD. We focused on gathering qualitative feedback, analysing technology acceptance through questionnaires, interviews, and recorded participants comments during the experiment. The collected data helped identify areas of the HUD system that require enhancement to improve usability and overall user experience. Our analysis includes a discussion on the qualitative feedback, emphasizing the role of user familiarity with the system in shaping their experience. This insight is crucial for refining the HUD's design to better meet the operational needs of soldiers in complex urban scenarios.

1.0 INTRODUCTION

Augmented Reality (AR) superimposes or integrates virtual content into real world environment [1]. Superimposed virtual elements are often unrealistic and they are displayed over real world objects with disregard to realistic overlay. Integrated virtual elements tend to be more realistic and need increased tracing and positioning to truly integrate them into the real world. XR (often used as an abbreviation for extended reality) is frequently used as an umbrella term for a variety of distinct concepts – most prominently AR and Virtual Reality (VR). The terminology used for presented terms differs in various research and commercial publications, which can lead to confusion. Therefore, we use a very detailed classification of different aspects of AR and VR given in [2]**.** Specifically, levels of local presence of virtual elements are used for categorization of AR services on a continuum between Assisted Reality and Mixed Reality as presented in Figure 1-1 [2]**.** Our focus is design of AR interfaces in geo-located AR displayed on wearable AR technologies encompassing different levels of local presence on the defined continuum. Geo-located AR merges AR and Geographic Information System (GIS), enabling the placement of virtual elements in user's surroundings, anchored to a specific geographic location. This means that virtual elements are visible to users only when they are in specific physical locations. For placing virtual elements, Geo-located AR relays on Global Positioning System (GPS) and other spatial information, like plane detection or data about the environment from GIS [3].

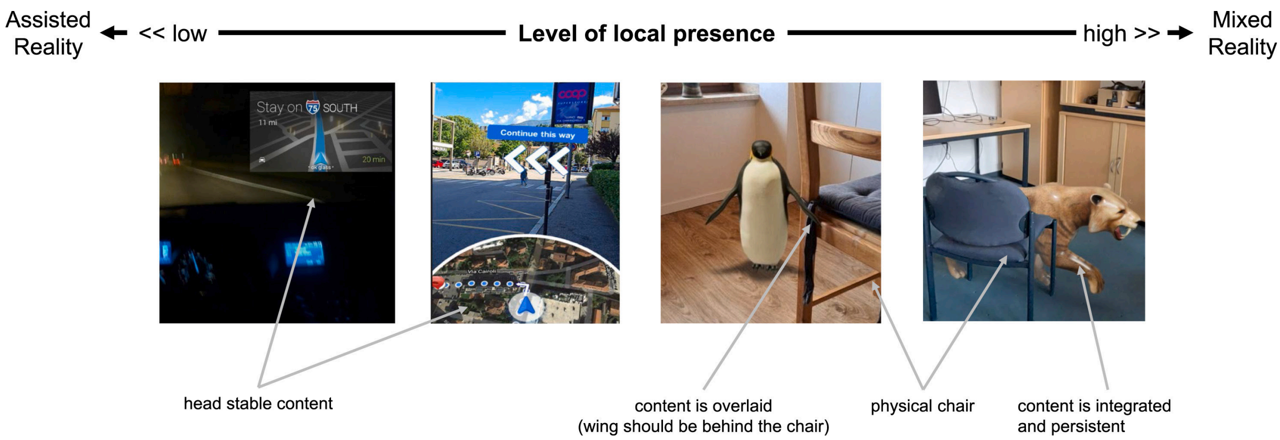

Figure 1-1: Four examples on the Assisted-Mixed-Reality Continuum (taken from [2]).

This type of AR usually integrates virtual elements into the real world but doesn't exclude superimposed elements. User's environment serves as a platform for providing virtual information, which in turn brings about new obstacles and challenges for interface design.

Interface determines how information is displayed and how the user interacts with it, that is why it represents the interaction point between the user and the system. Interfaces that adhere to design guidelines and patterns improve usability, allowing for more intuitive and effective user interactions, while ensuring system functionalities are accessible and easy to use [4]. AR must enable users to manipulate virtual elements while not disrupting their interaction with the real world. Given that AR either superimposes or integrates virtual elements, the presentation of information in AR is vital and requires a certain degree of abstraction.

Wearable AR are AR devices that users wear on their body, providing AR without the need to hold the device. These devices are mostly known as head-mounted displays (HMD). Head-up displays (HUDs) are interfaces for HMDs that provide users with information within their forward field of view (FFV). In AR, HUDs provide the user with the necessary information in their FFV without obscuring the real world, hence enabling user uninterrupted interaction with the real world. In combination with different input interactions, such as eye tracing or voice control, HUDs can provide the user with hands-free interaction with the system.

In this paper we will present a study which evaluates a geo-located AR user interface comprising UI elements from different points on the AR continuum. We create a video game for testing the developed interface. Our first iteration user testing was done using a video game as a simulator, while in future iterations we aim to use wearable AR devices.

2.0 RELATED WORK

According to Dey et al. [5], only 24 out of 291 research papers published between 2005 and 2014 focused on navigation, while just 28 of 369 user studies highlighted navigation as primary focus. There is a lack of AR research in military domain; Tezer et al. [6] found only 7 out of 1008 papers to focus on military.

Information visualization is a core challenge in AR interfaces, as combining virtual and real-world elements introduces novel issues that developers and designers haven't encountered previously [7]. Information loss is a big concern, therefore certain level of abstraction, adequate visualization and distribution in user's FFV is required. Hardware limitations of AR devices, such as restricted outdoor use and absence of features like occlusion, have prevented broader adoption, resulting in lack of models and guidelines for AR interface design [8][9]. Chen [8] also proposes general design principles for AR interfaces, although that is rather rare [4]. There is no standard for assessing the Quality of Experience (QoE) in AR systems, but rather general evaluation methods without precise guidelines [10][8]. Researchers combine different subjective (e.g. questionnaires) and objective metrics (e.g. performance) for QoE assessment. Gutiérrez et al. [11] identified an absence of consensus on which AR elements should be evaluated for QoE, contributing to a lack of standard assessment methods. However, while they classify evaluation attributes, they don't propose methods for assessing them.

Vasquez et al. [12] did a user study with both military personnel and civilians on Map in the Sky novel concept for navigation. They collected qualitative data from participants with an interview and measures on task performance and cognitive workload which were described in another paper. From the qualitative data they identified 6 main themes about map in the sky features. They found that the Map in the Sky has some advantages over traditional map displays, but also room for improvement.

Our research differs from previous ones as it focuses on military domain, encompassing UI elements from different points on the AR continuum. We aim to create guidelines which can drive AR interface design incorporating contexts of different uses (e.g., civilian, military, and entertainment) even for hardware functionalities of wearable AR devices which are yet to come.

2.1 Research Motivation and Importance

It is very important to establish design guidelines as well as QoE evaluation methods due to the complexity of displaying and distributing information in the user's forward field of view. Although AR technology is not yet fully developed, it is necessary to create guidelines, models and assessment methods for AR systems to prepare for the future AR technology features. By testing the proposed systems before the fully developed system functionalities, it shortens the total development time and reduces future problems that may arise.

2.2 Research Objectives

Our goals are to find a model or guidelines for the design of AR interfaces. We assume that different groups have different needs, therefore the design of AR interfaces should be adapted to each groups needs. With this research we intend to identify military groups needs and how they affect the interface design. Our research questions are:

- **RQ1** Are there differences in the tested metric between military and civilian groups?
- **RQ2** Does participants' age affect their affinity for technology interaction score?
- **RQ3** Will participants with higher affinity for technology interaction evaluate the system as more usable?
- **RQ4** Will participants with higher affinity for technology interaction be more satisfied with using interface elements?
- **RQ5** Will participants with higher ability to navigate perceive navigation elements as more useful?
- **RQ6** Will participants with higher ability to navigate be more satisfied while using the navigation elements.

3.0 METHODOLOGY

In this section we first describe the conceptualized interface elements. Following that, the study procedure is explained in detail, from participants to experimental design and data analysis.

3.1 Interface Elements

Conceptualized interface elements are divided into two groups: static information and spatial information. Firstly, it was considered whether the information has a spatial component, and then the importance of visibility and accessibility to the user. Static information is placed on the HUD and is always visible, while information with a spatial component is placed in the real world. To simplify the interaction with certain elements, the tilt of the user's head is used as an interaction system.

Spatial information is divided into three elements: Map in the Sky, navigational arrows, and tags for objects of interest. Map in the Sky is a previously researched concept for navigation [13]. The goal of the concept is to utilize the typically empty space in the sky, because it rarely displays important information for the user. Another benefit is that with this way of visualizing the map the user's view of the ground-level features isn't obstructed by AR navigation methods. Digital map of the user's surroundings is placed in the sky, only visible when the users head is tilted up towards the sky. The map is a mirror image of the user's surroundings and is perceived by the user as in same scale and with same distance as corresponding real-world features. Natural orientation was considered, so if the user perceives that they need to turn right on the map, this will correspond exactly to the real world. Map is offset in front of the user and rotates with them, so that the relevant area is always in front of the user [14]. User's position and positions of friendly and enemy combatants are depicted on the map with tags that follow their movement. Therefore, the map shows user’s surroundings and enemy and friendly combatants’ positions and movement. Navigational arrows are a separate navigational system and are not connected to the map. Arrows are placed on the ground and navigate the user from start to finish, only visible when the users head is tilted down towards the ground. Arrows transparency changes with user’s downward head tilt, making the arrows less transparent. Their size, placement and transparency assure that the arrows don’t obstruct the user's view and detection of obstacles. Tags for enemies and friendly combatants visualize their positions to the user. Their design is simple, two shapes of different colours that are always visible and are not connected to the user's head tilt. Tags follow the movements of combatants to accurately represent their position. The tags visible through other objects, e.g. buildings and fences. With this implementation the user can perceive the positions of persons and objects of interest even though their physical representations are obscured by physical objects in the environment.

3.2 Participants

We recruited 65 participants and divided them into 2 groups. The first group consisted of 31 participants (14 male, 17 female) who had no experience with military combat or exercises, so we called this group Civilians. The second group consisted of 34 participants (33 male, 1 female) who had previous military combat or exercises experience, and are not active military personnel anymore, so we called this group Military. The average age of all 65 participants is 44.6 years, the average age of civilian participants is 32.9 while the average age of military participants was 55.4.

3.3 Apparatus

The video game was created using Unity 2022.3.10f1 version and the game was built into a single application for Windows operating system. The project ran on Intel Core i7-13650 HX with 2.60 GHz core and a graphics card NVIDIA GeForce RTX 4050 on a 15.6-inch FHD (1920x1080) display. Users interacted with the video game via built in keyboard and Fantech Cruiser WG11 wireless 2.4 GHz pro-gaming mouse. The laptop was plugged in an electrical outlet and always in performance mode.

3.4 Video Game Simulation Design

Video game simulation was designed as a First-Person View (FPV) shooter game where the user controls the player with mouse and keyboard. We chose the FPV approach to provide the user with the same perspective as the player, enabling us to simulate what the player (soldier) would see through AR glasses with a HUD.

Two scenarios were created for the video game, a simple one as a tutorial and a more complex one for the game. The tutorial was made for the participants to familiarize themselves with the interface, movement and shooting mechanics. The game was made for testing the interface concept by playing through a specific scenario. The player is placed in an urban area, surrounded with buildings. In the game, the player has six friendly combatants and many enemies. The enemies are static at first and move when they spot the user. Standard controls and mechanisms from FPS games are employed.

The map is positioned in the sky and is hidden by objects, e.g. buildings and trees. Visibility of the map is dependent on user's head tilt. Level of details on the map is equivalent to a satellite image with simple 2D tags for the destination, user's, enemy and friendly combatants' positions. On the map the user is represented as a yellow arrowhead, while the enemies and friendly combatants are represented by red and green circles respectively. The destination is represented by two yellow rings. The map is depicted on the far-left image in Figure 1-2. Navigation arrows are a yellow 2D arrows that are placed in the middle of the road densely spaced. They are placed on all routs to ensure arrival to destination. Arrows are depicted in the middle image of Figure 1-2. Tags are 2D shapes with different colours for each category they are representing. For people, the tags are positioned above their heads with fixed size, while the destination has its own tag that is positioned above it. Enemies are represented by red triangles that are pointing downwards, while the friendly combatants are represented by green circles. The destination is represented by a yellow triangle that is pointing downwards. Since the tags are positioned above their heads, they change in size depending on the distance from the user. As mentioned earlier, the tags are visible through other objects and there is no cut of distance set, so the tag is visible to the user no matter how far away the tag is. Tags are displayed on the far-right image of Figure 1-2.

Figure 1-2 Map in the Sky is shown on the far-left image, navigational arrows are shown in the middle image and tags for enemy and friendly combatants are shown on the far-right image.

Users goal in the game is to reach the destination and stay in the destination zone for 10 seconds. There is no time constraint on the participants to reach the goal, nor are there any restrictions on the preferred path or number of combatants that must survive. Users should use all available AR elements to reach the destination.

3.5 Procedure

The study had multiple steps the participants had to go through that is depicted in Figure 1-3. Participants were informed of the research purposes, and they signed the consent form. After that the participants were informed of the study procedure and their task, and their anonymity was guaranteed.

Firstly, the participants filled out a general information questionnaire that had three distinctive portions. First portion consisted of questions for collecting demographic data, e.g. age, gender, level of education etc. With the targeted group of military veterans, information about their military rank was requested. They were further asked about their level of experience with video games and AR. The navigation portion of questionnaire consisted of 5 questions where the participants had to self-evaluate their navigational and orientation skills. The questions were formed as a Likert scale from 1 to 5, where 1 represented strongly disagree and 5 represented strongly agree. The third portion was the Affinity for Technology Interaction

(ATI) questionnaire [15]. The navigational questionnaire was created to study if there is a corelation between persons perceived navigation skills and scores of navigation elements from our interface. ATI scale was chosen to determine if there is a connection between the participants affinity for technology and the acceptance of our proposed interface.

After finishing the general information questionnaire, the participants open the video game and go through instructions for controlling the player, a short reminder of the AR elements positions and in game goals. They open the tutorial to practice controlling the player and to familiarize themselves with the AR elements. When the participants were ready, they would start the game. If the participant lost their life in the game, they could star over from the beginning.

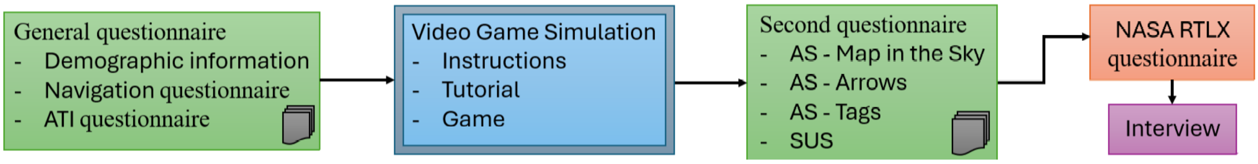

Figure 1-3 Study procedure.

At the beginning the participants were told that the experimenter will use the Think Aloud Protocol (TAP) and take note of their thoughts while they use the game if they decide to vocalize them. Usually when TAP is utilized, the experimenter encourages the participants to vocalize their thoughts, but we decided to only take notes if the participants had the natural urge to vocalize their opinions while playing the game.

With the game finished, the participants would start the second questionnaire that had four parts. The first three parts was the Acceptance Scale (AS) questionnaire for each AR element; Map in the sky, navigational arrows and tags for enemy and friendly combatants [16]. This questionnaire determines two dimensions: usefulness scale and satisfying scale. The fourth part of the questionnaire was the System Usability Scale (SUS) of the whole interface [17]. AS was chosen for each AR element because it determines two dimensions but is relatively short and simple, so we can determine how much was each element useful and satisfying to participants. SUS questionnaire was chosen to determine the overall usability of the interface.

NASA Raw Task Load Index (NASA RTLX) measures workload the participants perceives while doing a task [18]. The questionnaire consists of two parts, determining the source of the load and magnitude of the load. We decided to not include the weights determination for the first iteration, but rather give participants the raw version and let them rate each of 6 factors from 0 to 100.

The last part of the study procedure was an interview that consisted of 4 general questions and a question for each element of interest. Participants were asked if they liked the concept, which of the proposed elements would they single out as either positive or negative and what would they change about them. For each element the participants were asked if they liked it, if it was useful to them and if they would change it. This was followed up with some sub-questions about each element, e.g. colour of arrows, size etc.

3.6 Data Analysis

For data analysis we used excel tables and a program in Python language for complete data analysis and graph plotting. Final navigational score went from 1 to 5. Final ATI score goes from 1 to 6. Since AS questionnaire determines two dimensions: usefulness scale and satisfying scale, we had to calculate both scores. The resulting scores go from -2 to 2. Resulting SUS scores go from 0 to 100. NASA RTLX calculates the average individual score, average overall score and average overall score for each factor.

For both participants' gender and age, we calculated the average of all participants and each group. For navigational questionnaire, ATI, AS and SUS we calculated the mean, median and standard deviation for all participants and each group. For each interview question and variable within the question we calculated the distribution. Pearson correlation, Spearman's rank correlation, Kendall's Tau correlation and Kruskal-Wallis test, were implemented for analysing different types of corelation between data.

4.0 RESULTS AND DISCUSSION

4.1 Questionnaires Results

Participants perception of their navigational skills is rather high (mean=4.4154, stddev=0.6557). Military group has a higher score (mean=4.4706) than the civilian group (mean=4.3548).

Participants' ATI scores are in the upper half (mean=3.7966, stddev=1.1173), indicating a higher affinity for technology (see graph a) in Figure 1-4). The military group's ATI score (mean=3.4346, stddev=1.0961) is significantly lower than the civilian group's (mean=4.1935, stddev=0.9994). This difference may be due to the military group's older age and their military experience from around 30 years ago, during which technology has advanced rapidly.

SUS scores are very similar across groups and are all within 96 to 100 percentile range giving them a grade of A+, as visible on graph b) in Figure 1-3. Military group (mean=86.0294, stddev=14.9154) has a lower score than the civilian group (mean=89.1129, stddev=10.6360). Even with standard deviation, all groups have an acceptable score above 51.6 and are above the 50th percentile [19].

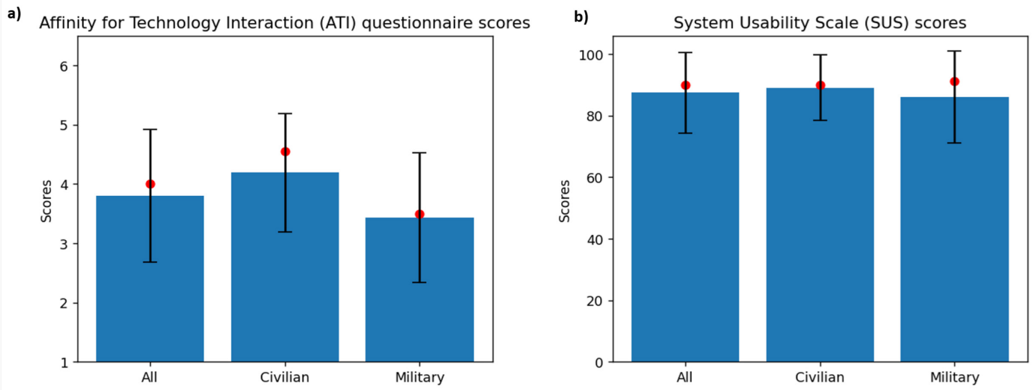

Figure 1-4: Graph a) depicts ATI scores, graph b) depicts SUS scores.

Map in the Sky element has a positive AS score across all groups. The usefulness scale shows a slight positive effect for all participants (mean=0.7723, stddev=1.0227), while the satisfaction scale is lower (mean=0.6192, stddev=0.9865) as visible on graph a) in Figure 1-5. Civilians rated the map more useful (mean=0.8194) than the military (mean=0.7294), but the military found it more satisfying to use (mean=0.6324 vs. 0.6048 for civilians). High standard deviations across both dimensions suggest the need for improvements, likely due to participants' unfamiliarity with this new map concept.

Navigation arrows received a much more positive AS score across all groups. The usefulness scale for all participants (mean=1.5569, stddev=0.7313) shows a highly positive effect, with a slightly lower satisfaction score (mean=1.4577, stddev=0.6872) as visible on graph b) in Figure 1-5. The military rated the arrows more useful (mean=1.7294) and satisfying (mean=1.5809) than the civilian group (usefulness: mean=1.3677, satisfaction: mean=1.3226). Despite high standard deviations, the results remain positive.

Tags for enemy and friendly combatants received the highest and most positive AS scores across both dimensions. The usefulness scale is highly positive for all participants (mean=1.7969, stddev=0.3961), with slightly lower satisfaction (mean=1.5807, stddev=0.5786) as visible on graph c) in Figure 1-5. The military group rated the tags more useful (mean=1.8471) and satisfying (mean=1.625) than civilians (usefulness: mean=1.7419, satisfaction: mean=1.5323). While usefulness scores had low variability (stddev < 0.5), satisfaction scores showed more variance, indicating some room for improvement. Despite this, the overall results are highly positive.

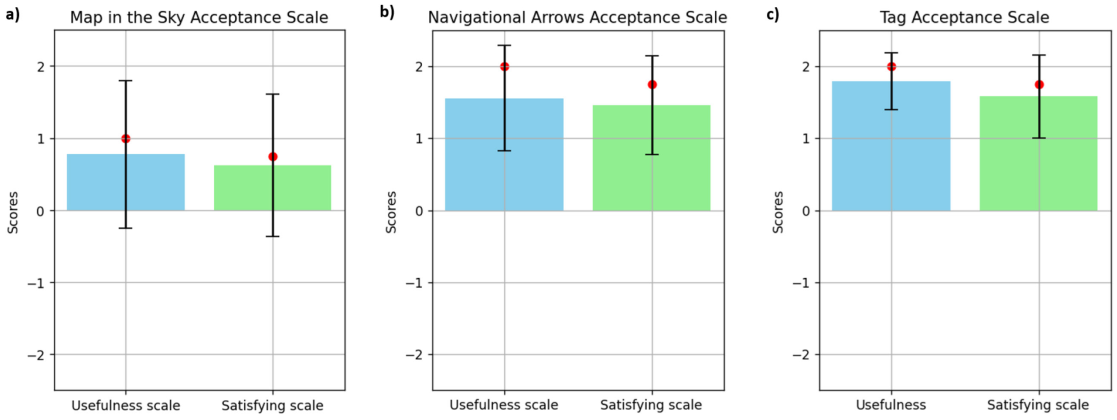

Figure 1-5: Graph a) shows AS score for Map in the Sky. Graph b) shows AS score for Navigational arrows. Graph c) shows AS score for Tags for enemy and friendly combatants.

4.2 Interview Results

In response to the first question, 97.06% of military and 100% of civilian participants liked the AR interface. Both groups highlighted the navigational arrows as the most positive element (military 42.55%, civilians 35.90%). The military's second choice was the tags (38.29%), while civilians preferred the map (33.33%). Most participants (military 61.76%, civilians 54.83%) couldn't identify a negative element, with the map being the most common critique. When asked about changes, most military participants (58.82%) would change nothing, while 48.38% of civilians would modify the map. For the military group, 32.35% selected the map as the second element for change, while 35.48% of civilians did not want to change anything. Interview results for each element are shown in graphs. Graph a) in Figure 1-7 represents results for Map in the sky, while graph a) in Figure 1-6 represents results for Navigational Arrows. Graph b) in Figure 1-6 represents results for Tags for enemy and friendly combatants, while graph b) in Figure 1-7 shows results for all three elements when the question had 3 possible answers.

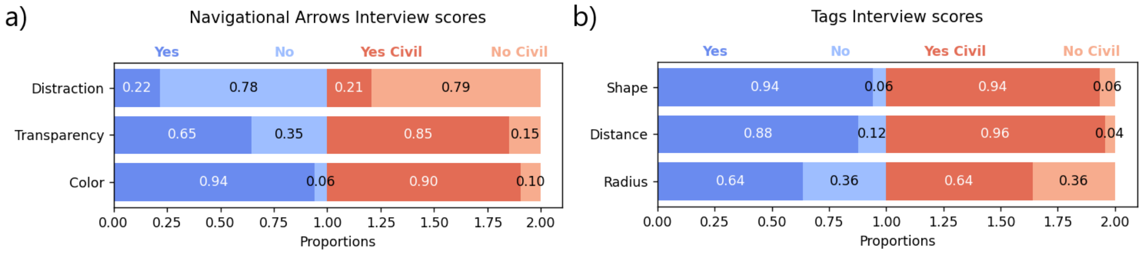

Figure 1-6: Graph a) is interview results for arrows. Grap b) is interview results for tags.

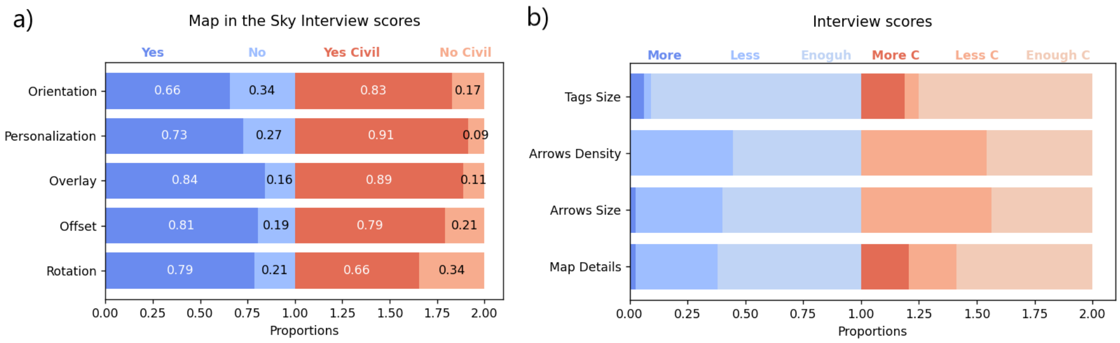

Figure 1-7: Graph a) shows interview results for the map. Graph b) shows interview results for all elements when the answer had 3 options

We performed Chi-Square Test for each feature of every element to determine if there are any significant differences between groups and the only significant difference we found was for details on the map. This result could be because our groups have different needs, therefore they want different levels of details on the map. Also, we cannot rule out age as a possible factor of this result. To **answer RQ1, there is one significant difference between groups.**

4.3 Correlations

Three correlations were found for the relationship between the age of all participants and their ATI scores **answers RQ2 positively**. A Pearson correlation value of -0.35 was found, which means that there is a weak linear negative relationship between the age of the participants and the ATI score, as visible on the graph a) in Figure 1-8. Since it is a weak correlation, it is not surprising that Spearman rank correlation with a value of -0.31 and Kendall tau correlation with a value of -0.21 were also found. Both correlations indicate a weak negative nonlinear relationship. Because the correlations are weak, we can see on the graphs that the linear function as well as the monotonic function do not fully represent the dispersion of the data, but that there is a link between the age of the participants and the ATI results.

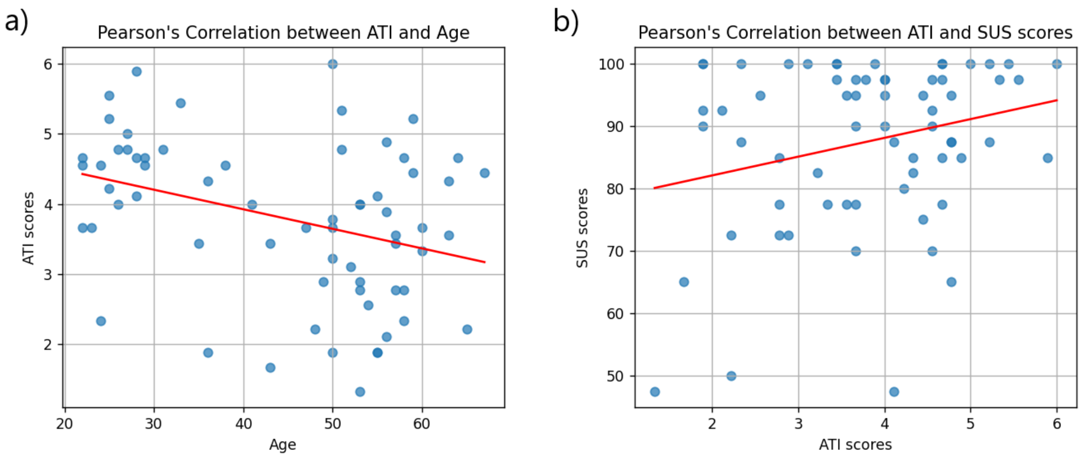

Figure 1-8: graph a) shows Pearson's correlation between ATI score and Age. Graph b) shows Pearson's correlation between ATI and SUS scores.

Linear correlation between ATI scores and SUS scores for all participants was found, with Pearson's correlation value of 0.26, indicating a weak positive relationship, as visible on the graph b) in Figure 1-8 thus

answering RQ3 positively. This would mean that users with more affinity towards technology interactions would find the system more usable.

There is a Kendall's tau correlation between the navigational questionnaire and usefulness scale for navigational arrows with value of 0.20, indicating a very low positive association. Another Kendall's tau correlation was found between the navigational questionnaire and satisfying scale for navigational arrows with value of 0.23, also indicating a very low positive association. This means that both dimensions of the AS questionnaire slightly depend on the user's perception on their navigation skills, higher navigational questionnaire score indicates higher AS questionnaire scores. For the relationship between the navigational questionnaire and satisfying scale for navigational arrows we also found Spearman's rank correlation with value 0.29, indicating a low positive association. Both statistics confirm a low positive relationship, meaning that as one variable increases, the other tends to increase. There is also a Pearson's correlation between the navigational questionnaire and satisfying scale for navigational arrows for the military group with a value of 0.35, indicating a weak to moderate linear relationship, which corresponds with earlier mentioned positive correlations. These correlations give us a **positive answer to both RQ5 and RQ6**.

Pearson correlation and Kendall's tau correlation were found between the military group ATI scores and satisfying scale scores for the Map in the Sky, with values of 0.38 and 0.25 respectively. This indicates a weak linear positive relationship and a weak nonlinear positive relationship between the data. Finally, a Pearson correlation with value of 0.35 was found was between the military group ATI scores and satisfying scale scores for the navigational arrows. This indicates a weak linear positive relationship between the ATI and AS scores, meaning that military participants with higher ATI scores found the navigational arrows and map satisfying to use, answering the **RQ4 positively**.

5.0 CONCLUSION AND FUTURE WORK

The results indicate that the proposed AR interface design was generally accepted by participants, though larger standard deviations suggest room for improvement. The Map in the Sky received the lowest scores, likely due to its novelty. Many participants expressed that with more practice, their performance and perception of usefulness could improve. Interviews revealed the concept's potential, but highlighted the need to optimize the map's angle, height, and level of detail. Participants noted that tall buildings obstructed views and liked the idea of overlaying the map, which should be included in the next iteration, along with a personalization menu. While navigational arrows were better received than the map, they also require optimization in size and placement; some suggested a single, rotating arrow indicating direction. Tags for combatants received positive feedback, with only minor revisions needed. Opinions on the radius feature were mixed, warranting further testing. Suggestions for varied tag shapes and colours to represent different threats should also be explored. All elements scored lower in satisfaction than usefulness and many participants expressed their dislike of the questions in the AS questionnaire, indicating a need for a revised tool to better assess user satisfaction and the effectiveness of AR features.

Linear and nonlinear correlations between age and ATI scores suggest that age does affect a person's ATI score, but the low correlations suggest a more complex relationship between variables and there might be more parameters that affect the ATI score besides age. We found a low correlation between ATI and SUS scores, meaning that users with higher ATI scores have higher SUS scores and are more likely to say that a system is usable. The main reason that this is the case could be the fact that people who have greater affinity for technology interaction also find the system more usable. This could be because they have more experience and don't find new systems complicated to use. Another reason could be the age factor; therefore, this relationship should be investigated further. A weak correlation was also found between users' perception of their ability to navigate and user satisfaction with navigational elements. It is necessary to examine further the relationship between the user's ability to navigate and to use the proposed navigational elements.

Given that military participants are no longer active soldiers, but veterans and some trends that appear in their results cannot be adequately compared with civilians, it would be useful to examine currently active soldiers or even those who are learning to be soldiers to see how much age and experience influence the perception of military participants. With this new group of participants, we could get a much better overview of our system.

Our results were much more positive than Vasquez et al., but we confirmed their conclusions on novelty of the concept affecting the usefulness of the map, with participants saying they would perform better if they had more practice. We also confirmed that some participants have trouble orienting on the map, find the map's rotation confusing and positioning in the sky to be at a too high angle so they had to tilt their head to far up. Participants also wanted the map to only show on demand, same as participants in Vasquez's study. Finally, we confirmed that participants want the ability to personalize the map to their needs.

6.0 ACKNOWLEDGEMENTS

Research was supported by the Croatian National Recovery and Resilience Plan (NPOO) under the project Research and Development of Multiple Innovative Products, Services and Business Models Aimed at Strengthening Sustainable Tourism and the Green and Digital Transition of Tourism (ROBOCAMP), with grant number NPOO.C1.6.R1-I2.01, and XR Communication and Interaction Through a Dynamically Updated Digital Twin of a Smart Space – DIGIPHY with grant number NPOO.C3.2.R3-I1.04.00700.

7.0 REFERENCES

[1] b24 Billinghurst, M., Clark, A., & Lee, G. (2015). "A survey of augmented reality." Foundations and Trends® in Human–Computer Interaction, 8(2-3), 73-272.

[2] b25 Rauschnabel, P. A., Felix, R., Hinsch, C., Shahab, H., & Alt, F. (2022). "What is XR? Towards a framework for augmented and virtual reality." Computers in human behavior, 133, 107289.

[3] b4 Safari Bazargani, J., Zafari, M., Sadeghi-Niaraki, A., & Choi, S. M. (2022). "A survey of GIS and AR integration: applications." Sustainability, 14(16), 10134.

[4] b6 Kim, K., Billinghurst, M., Bruder, G., Duh, H. B. L., & Welch, G. F. (2018). "Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008–2017)." IEEE transactions on visualization and computer graphics, 24(11), 2947-2962.

[5] b15 Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. E. (2018). "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014." Frontiers in Robotics and AI, 5, 37.

[6] b14 Tezer, M., Yıldız, E., Masalimova, A., Fatkhutdinova, A., Zheltukhina, M., & Khairullina, E. (2019). "Trends of augmented reality applications and research throughout the world: Meta-analysis of theses, articles and papers between 2001-2019 years." International Journal of Emerging Technologies in Learning (iJET), 14(22), 154-174.

[7] b11 Zollmann, S., Langlotz, T., Grasset, R., Lo, W. H., Mori, S., & Regenbrecht, H. (2020). "Visualization techniques in augmented reality: A taxonomy, methods and patterns." IEEE transactions on visualization and computer graphics, 27(9), 3808-3825.

[8] b12 Chen, S. (2024). "Augmented Reality User Interfaces: Analyzing Design Principles and Evaluation Methods for Augmented Reality (AR) User Interfaces to Enhance User Interaction and Experience." Human-Computer Interaction Perspectives, 4(1), 15-27.

[9] b13 Lovreglio, R., & Kinateder, M. (2020). “Augmented reality for pedestrian evacuation research: promises and limitations.” Safety science, 128, 104750.

[10] b23 Pérez, P., Gonzalez-Sosa, E., Gutiérrez, J., & García, N. (2022). “Emerging immersive communication systems: overview, taxonomy, and good practices for QoE assessment.” Frontiers in Signal Processing, 2, 917684.

[11] b38 Gutiérrez, L. E., Betts, M. M., Wightman, P., Salazar, A., Jabba, D., & Nieto, W. (2022). “Characterization of Quality Attributes to Evaluate the User Experience in Augmented Reality.” IEEE Access, 10, 112639-112656.

[12] R3 Vasquez, H. M., Reiner, A. J., Jamieson, G. A., & Hollands, J. G. (2020, December). “A Mirror-in-the-Sky Navigation Aid: Summary of Qualitative Feedback from Soldier and Civilian Users.” In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 64, No. 1, pp. 2030-2034). Sage CA: Los Angeles, CA: SAGE Publications.

[13] R1 Kapler, T., King, R., & Segura, D. (2019, October). ““SkyMap”: World-Scale Immersive Spatial Display”. In Symposium on Spatial User Interaction (pp. 1-3).

[14] R2 Reiner, A. J., Hollands, J. G., Jamieson, G. A., & Boustila, S. (2020). “A mirror in the sky: assessment of an augmented reality method for depicting navigational information.” Ergonomics, 63(5), 548-562.

[15] R4 T. Franke, T., Attig, C., & Wessel, D. (2019). “A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale.” International Journal of Human–Computer Interaction, 35(6), 456-467.

[16] R7 Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). “A simple procedure for the assessment of acceptance of advanced transport telematics.” Transportation Research Part C: Emerging Technologies, 5(1), 1-10.

[17] R5 Brooke, J. (1996). “SUS: A quick and dirty usability scale.” Usability Evaluation in Industry.

[18] R6 Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). “Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary.” Advances in ind. ergonomics and safety, 1, 481-485.

[19] R8 Lewis, J. R., & Sauro, J. (2018). “Item benchmarks for the system usability scale.” Journal of Usability Studies, 13(3).