

Augmented Reality-Based Human Memory Enhancement Using Artificial Intelligence

Zhanat Makhataeva, Tolegen Akhmetov, and Huseyin Atakan Varol, *Senior Member, IEEE*

Abstract—Augmented reality (AR) offers novel ways to design, curate, and deliver information to users by integrating virtual, computer-generated objects into a real-world environment. This work presents an AR-based human memory augmentation system that uses computer vision (CV) and artificial intelligence (AI) to replace the internal mental representation of objects in the environment with an external augmented representation. The system consists of two components: (1) an AR headset and (2) a computing station. The AR headset runs an application that senses the indoor environment, sends data to the computing station for processing, receives the processed data, and updates the external representation of objects using a virtual 3D object projected into the real environment in front of the user’s eyes. The computing station performs computer vision-based indoor environment self-localization, object detection, and object-to-location binding using first-person view (FPV) data received from the AR headset. We designed a behavioral study to evaluate the usability of the system. In a pilot study with 26 participants (12 females and 14 males), we investigated human performance in an experimental task that involved remembering the positions of objects in a physical space and displaying the positions of the learned objects on the two-dimensional (2D) map of the space. We conducted the studies under two conditions—that is, with and without using the AR system. We investigated the usability of the system, subjective workload, and performance variables under both conditions. The results showed that the AR-based augmentation of the mental representation of objects indoors reduced cognitive load and increased performance accuracy.

Index Terms—Human memory augmentation, augmented reality (AR), computer vision (CV), artificial intelligence (AI), spatial cognition, object-location memory, navigation, cognitive load, performance, system usability.

I. INTRODUCTION

AUGMENTED reality (AR) displays are emerging as a next-generation interactive technology that enriches people’s three-dimensional (3D) visual experiences in the real-world environment [1]. AR systems introduce people to a supplemented interactive environment that can combine real and computer-generated objects in real space [2]. In AR, virtual objects appear to coexist with real objects in specific locations in the real world [3]. Virtual reality (VR), another immersive technology, introduces users to a fully computer-generated virtual world that is independent of the real world [4]. There is much recent work in the literature that addresses the

perspectives of augmented visualization, sensing, and interactivity introduced by AR displays to support human cognitive processes, including memory [5], spatial cognition [6], and navigation [7]. Thanks to spatial and environmental cognition, people acquire, store, recall, and decode information about relative locations and create cognitive maps that are internal representations of space and the corresponding attributive values and meanings [8]. According to the authors, spatial cognition refers to internalized reflections and reconstructions of the interacted space in thought. In [9], Burgess reviewed advances in understanding spatial cognition from a memory perspective. The author described egocentric and allocentric representations of location and self-motion involved in spatial memory, navigation, and imagery. Egocentric representations relate to sensory information and are associated with coordinate frames of specific receptors, for example, retinotopic for vision, head-centered for hearing, and body-centered for actions. Allocentric representations are more abstract and focus on landmarks in the environment. Human abilities to create cognitive maps, spatial learning, and navigation are closely linked to memory.

Human memory stores and retrieves the information we acquire through our senses [10]. As a complex system, human memory relies on components, such as sensory, short-term, and long-term memory [10]. Sensory memory refers to a specific sensory modality (e.g., visual and auditory memory). Working memory (also called short-term memory) has a limited capacity and simultaneously stores and processes goal-relevant information for the small period of time (i.e., a brief period of up to a few seconds). Long-term memory has an unlimited capacity and can store large amounts of information indefinitely. According to the authors in [11], [12], working memory supports human thinking and cognitive processes by acting as an interface between human perception, long-term memory, and action.

The working memory system comprises a central executive system and two storage systems (i.e., the phonological loop for verbal data chunks and the visuospatial sketchpad for visual-spatial data chunks) [12]. In [13], Smith et al. described three working memory systems that process spatial, object, and verbal information. The authors related spatial and object coding phenomena to two distinct visual processes that encode “what” and “where” information. Postma et al. [14] examined the neurocognition of object-location memory. The authors found that object-location memory is functionally divided into three processing mechanisms—namely, object processing, spatial-location processing, and object-to-location binding. According to O’regan [15], visual retention is mainly redundant and

Manuscript received May, 2022; revised 2022; accepted 2023. Date of publication 2023; date of current version 2023. This article has supplementary downloadable material available at <https://github.com/IS2AI/ExoMem-AR-Memory>. (Corresponding author: Huseyin Atakan Varol.)

The authors are with the Institute of Smart Systems and Artificial Intelligence (ISSAI), Nazarbayev University, Astana 010000, Republic of Kazakhstan (e-mail: zhanat.mahataeva@nu.edu.kz; tolegen.akhmetov@nu.edu.kz; ahvarol@nu.edu.kz).

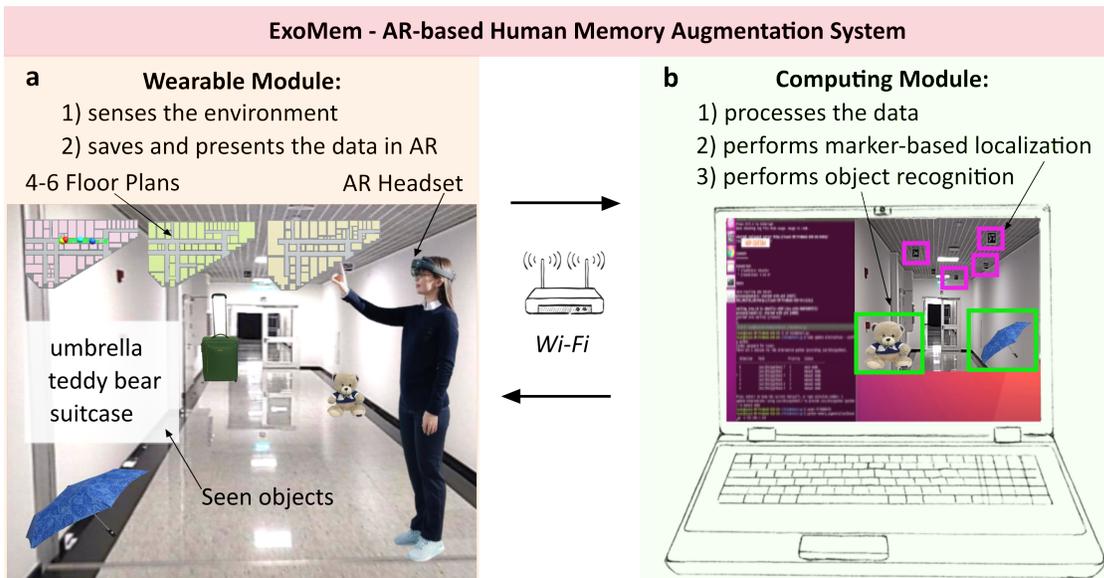


Fig. 1. Our AR-based human memory augmentation system (ExoMem) consists of wearable (a) and computing (b) modules communicating with each other wirelessly.

is followed over time by continuous generation of memory records. As Baddeley noted in [12], visual working memory and its verbal counterpart have a limited capacity (i.e., up to three or four objects).

The mystery behind the strict limits on how much data that can be kept in mind at once (3-5 meaningful items) was explored in [16], [17]. In [18], Sweller described recent advances in the cognitive load theory (CLT) and introduced a new line of research aimed at improving learning through an innovative design that takes into account the limits of internal working memory capacity. In [19], Van et al. presented several assumptions of the CLT related to memory, learning processes, the different types of cognitive load (i.e., intrinsic, extraneous, and germane), and the effects of design on learning [19].

Over the last decade, many studies have presented AR-enhanced systems for training and supporting cognitive skills [20]–[23]. However, there is limited work attempting to augment human memory. The best-known wearable application to augment human memory was SenseCam [24]. It was presented over a decade ago and used a wearable camera to record the user’s day by capturing images every 30 seconds. Among the recent applications dealing with human memory augmentation and AR, most, if not all, AR-based approaches have been presented to aid the memorization process or train specific types of memory [25]. In the literature, our work is the first to utilize an augmented environment and infrastructure-free solution for indoor user localization. We present a new approach to augment human memory based on AR visualizations and AI data processing. The AR system constructs the external augmented representation of objects in the indoor environment to replace internal mental representations.

We have named our human memory augmentation system “ExoMem” (see Fig. 1). The system consists of two main components: an AR headset and a computing station. The two components exchange data over a wireless network created by a router. The AR headset runs the Unity-based

application that senses the environment, sends image data to the computing station, and receives processed information about the environment. Based on the received data, the AR application creates and updates a virtual 3D object with the user’s current position, previous positions, and the locations of detected objects in the environment. All the information is incorporated into the virtual 3D object, which shows a 2D plan of the indoor environment projected in front of the user’s eyes in the real environment. Computer vision (CV) and artificial intelligence (AI) algorithms run on the computing station to perform indoor localization of the user and register the locations of objects in the environment. In a pilot study, 26 human participants (12 females and 14 males) used the AR system to create an external representation of ten objects placed on three floors of a building and later completed a computer-based test in which they displayed the locations of the previously learned objects on the 2D map of the building. Participants also completed the same procedure without the assistance of the AR system. They had to create an internal mental representation of the objects in the environment and complete the computer-based test using only their memory. We examined participants’ cognitive load and performance when they completed the object location memorization and map-pointing activities under two conditions—that is, with and without the AR system. We evaluated the usability of the system and investigated whether the gender of our participants affected the results of the behavioral study. Our work showed that the AR and AI-enhanced system can generate an augmented external representation of objects in physical space by interacting with the environment to replace the internal mental representations and support the human cognitive system.

The rest of the article is organized as follows: Section II reviews advances in cartography, CV, AI, and AR-based applications in navigation, spatial learning, and memory. Section III describes the software and hardware used in the system, the experimental setup, including object location memorization

and map-pointing activities, experimental procedure, and experimental measures. Section IV presents the study results with a thorough statistical analysis. Section V discusses the results, including an overview of the limitations of our study and suggestions for future research. Finally, Section VI concludes the paper.

II. RELATED WORK

Cartography deals with the art, science, and technology behind the creation and use of maps [26]. Key elements are the collection, acquisition, processing, and display of spatial data. Clarke et al. [27] summarized recent developments in cartographic research and divided them into different categories such as information visualization, cartographic data, spatial analysis, methods, and geographic information science. Yuan et al. [28] provided an overview of the relationship between cartography and information visualization. White et al. [29] presented the concept of remote sensing imagery and its role in information acquisition and reasoning in our natural environment [29]. According to the authors, technological advances introducing new sensing systems, new data types, new visualizations, new software systems for image processing and AI systems for scene and pattern recognition have shaped the state of the art in remote sensing and perception over the past two decades. The authors noted that technological advances have been accompanied by advances in the understanding of perceptual learning and reasoning in image analysis. In [30]–[32], the researchers explored the application of empirical and technological methods in the field of design and scientific visualization, including interactive maps and virtual reality environments supporting phenomena of digital interactivity. According to Roth et al. [31], the interactivity phenomena of interactive maps and visualizations require that the map user be involved in the creation of the representation, rather than just being a passive reader of the information. The authors pointed out that millions of people worldwide use interactive map services, such as Apple Maps, Google Maps, and OpenStreetMaps on a daily basis, indicating the high usability of such augmentative systems.

Emerging AR technologies enable the integration of volumetric virtual visualizations into the specified location in a 3D environment [5]. Few works relate AR to the graphical user interface used to interact with spatial data, also referred to as an augmented map [6]. In [33], the authors discussed the potential of augmented maps to enrich the functionality and performance of printed maps and to take user interaction with cartographic data to a new level. The interactivity and visualization capabilities of AR make a valuable contribution to supporting activities such as spatial orientation, spatial navigation, search processes, and decision-making [5]. According to Bobrich et al. [33], the usability of AR-enhanced augmentations is highly dependent on the user experience and the capabilities of a computer program. The authors noted that AR combines the latest inventions and capabilities of computer science and presents them to the user. For example, the integration of advanced computing and remote tracking systems in CV has led to advances in a wide range of intelligent

algorithms, such as simultaneous localization and mapping (SLAM), object detection and tracking, facial emotion and expression recognition, human action and activity recognition, hand gesture recognition, and head pose and gaze estimation [34]. Satellite-based geospatial positioning was utilized for outdoor navigation [35]. For indoor navigation, methods based on wireless signals (e.g., signals from Wi-Fi routers and Bluetooth beacons) and computer vision (e.g., features from images and fiducial markers) were developed [36].

In the last two decades, many researchers, engineers, and practitioners have proposed new approaches to navigation, tracking, and 3D positioning in AR. For example, a SLAM-based 3D positioning system was designed for an AR handheld system [37]. Rehman et al. [38] developed an AR-based indoor navigation system that used a pre-scanned 3D map of the surroundings (3D point clouds) to track the environment. Wearable cameras, gaze trackers, and CV algorithms have been used to recognize objects and actions in the environment so that video of a detected object and associated actions can be displayed based on user gaze [39]. In [40], Calle-Bustos et al. presented an AR application for supporting the user navigation indoors based on AR visual and auditory stimuli.

In addition, many papers have presented behavioral experiments using various AR-enhanced approaches to study visual-spatial working memory and visual perception. For example, Carbonell et al. [6] investigated how AR-based visualization of landmarks affects map reading during comprehension of geographic relief. Similarly, in their recent work, Keil et al. [5] studied whether AR-based holographic grids on the floor help people to form a better mental representation of the interacted space and support spatial learning. In [7], Chen et al. explored how AR-based instructions assist firefighters in wayfinding and decision-making during emergencies. Juan et al. [23] developed an AR-based task to assess the spatial short-term memory of preschool and primary school-aged children. In their recent work, Munoz et al. [41] presented a framework for developing an AR-based mobile application to assess human spatial short-term memory. In [42], Caluya et al. studied the impact of the spatial memory training framework on short- and long-term memory when conducted in AR and VR environments.

Furthermore, many works have analyzed AR-based visualization systems from the perspective of cognitive workload. For example, Atici et al. [20] examined how AR glasses affect the cognitive load of individuals working on the assembly line of an automotive company. During the behavioral experiment, workers completed the diffusion task (i.e., placing ten materials from the shelf into the appropriate carts) with and without the assistance of AR glasses. They reported lower cognitive load when they completed the task with the AR glasses. In [21], the authors evaluated an AR-assisted data collection system that supports the principles of spatial and temporal contiguity necessary in physics laboratory experiments. The authors found that the AR-based workflow reduced undergraduate students' cognitive load and helped them to avoid split attention while learning. In [22], the AR-based training system enabled junior nursing students to have an overview of internal structures and anatomical landmarks during clinical

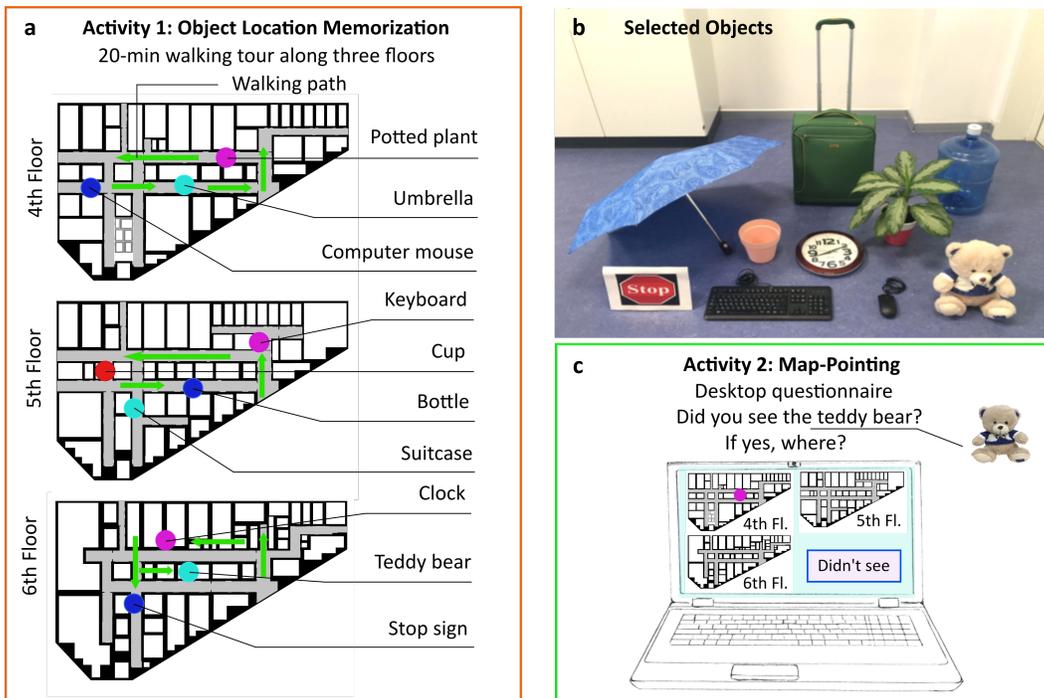


Fig. 2. **Experimental task with two activities and selected objects.** (a) Activity 1: Memorization of object locations placed along corridors during a 20-minute walking tour of three floors. (b) Ten objects were selected for the experimental task. (c) Activity 2: Showing the locations of the previously learned objects on the map during the computer-based test showing three 2D plans of floors 4-6 of the building.

skills training.

III. SYSTEM DESIGN AND EXPERIMENTAL SETUP

A. System Design

Our AR-based human memory augmentation system, “ExoMem”, is shown in Fig. 1. The first part of the system is a Microsoft HoloLens 2 mixed reality headset (see Fig. 1a). This AR headset is equipped with see-through holographic lenses, sensors for human and environment understanding, a holographic processing unit with 64-GB UFS 2.1 data storage, the Windows Holographic operating system, and lithium-ion batteries that allow 2-3 hours of active use. The open-source CV application programming interfaces and the manufacturer’s tools enable the wireless transmission of image data from the sensors to other computing devices. The second part of the AR system is an x64-based laptop computer running Ubuntu 16.04, with the Robot Operating System (ROS) installed (see Fig. 1b).

Wearable module: The main component of the wearable module of “ExoMem” is an FPV application running on the HoloLens 2 AR Goggles to sense the environment and communicate with the AI part of the system (see Fig. 1a). The Unity 3D 2018.4.28f1 game development engine was used to build and run the application on the Universal Windows Platform. The C# programming language was used in the codes to transmit the FPV data from the AR goggles to the laptop, receive the AI-processed data, and create an external representation of the objects in the indoor environment. The virtual 3D object was integrated with virtual 2D plans of three building floors and information about the user’s current position, previous positions, and locations of objects in the

environment. On the external representation, the application marked the path walked as green spheres and detected objects as spheres in other colors. The colored spheres on the virtual 3D object indicated the user’s position when the objects were first detected. The Unity project and source codes for sensing the indoor environment using the AR HoloLens 2 headset and creating the external representation of the objects in the environment are available in our GitHub repository¹ under the MIT license.

Computing module: The main software components of the “ExoMem” computing module are illustrated in Fig. 1b. For positioning the user in the research building of Nazarbayev University, we used the CV-based method for camera localization in indoor spaces, which was previously developed in [43]. The proposed localization method uses ArUco fiducial markers and the OpenCV library [44] to find the camera positions over time with respect to the registered markers in the building. The computing module of the system uses the ROS environment to receive FPV data from the HoloLens and run the user localization framework. The FPV data were also used to detect objects. For this purpose, we used the real-time object detector YOLO, developed in the Darknet framework [45]. The YOLO object detector is a deep end-to-end network that receives RGB image data as input and outputs the object labels and bounding box information as a string message. Specifically, in the AR part of the system, we used a pre-trained YOLO Version 3 object detection model trained on the COCO dataset with the default 80 object categories [46]. From these 80 categories, we selected ten objects for the experimental task. These ten objects were those that could be easily found in the

¹<https://github.com/IS2AI/ExoMem-AR-Memory>

laboratory environment (see Fig. 2b). The code running the YOLO object detector in our system was written in Python 3.7 and integrated into the ROS environment. All codes utilized for CV-based indoor user localization and object detection using FPV data from HoloLens 2 can be downloaded from our GitHub repository¹.

Data communication: The communication pipeline between the two modules (i.e., wearable and computing) was set up using a Wi-Fi network and TCP/IP network protocol via a portable router. The FPV application in the HoloLens sent images to the computing module at ten frames per second. To construct an external representation of objects in the environment, the AR goggles received string messages from the computing module, performed message parsing, and updated the virtual 3D object that represented the object and user location data processed by the AI part of the system. In the HoloLens, different color spheres were created on the augmented visualization about twice per second, showing the external representation of objects and the user's past and current positions in the environment.

B. Experimental Setup

We conducted two experimental sessions with each participant on two separate days with a two-week break between sessions. We scheduled a two-week break between sessions to allow participants sufficient time to forget the locations of the objects after the first session before coming to the second session. On both days (i.e., sessions), participants completed the same experimental task but under two conditions (i.e., with and without AR). To ensure that the order of conditions did not affect the results, we randomly divided our participants into two groups, taking care to maintain a gender balance between the groups. The first group completed the first session with the AR system and the second session without the AR system. The second group completed the same sessions in reverse order. We collected data three days a week, inviting two people to the experiment each day. We spent two weeks collecting the first set of data from all participants in the first group (6 females and 6 males). Then we invited the same individuals for the second session in the same order. Once we completed two data collection sessions for the first group of participants, we began data collection for the second group. We followed the same data collection procedure for the second group of participants and maintained the two-week break between the two sessions.

In both experimental sessions, we examined participants' cognitive workload. In each experimental session, participants completed two activities (e.g., memorizing object locations in the environment and pointing to the locations of learned objects on the 2D map of the environment). Three floors of a building on a university campus served as the experimental site. Each floor was 57 m by 90 m, and the length of the corridors on each floor was approximately 150 m.

Object location memorization activity (Activity 1): In the object location memorization activity, participants completed a 20-minute walking tour of the three floors of the building. They had to memorize the locations of ten objects (one at a time) that they saw along the way (see Fig. 2a). For this

purpose, we selected ten objects of different sizes and purposes that could be easily found in the laboratory environment (see Fig. 2b). The selected objects were placed along the corridors of the floors as shown in Fig. 2a. The positions of the objects were fixed at the beginning of the experimental study and were not changed thereafter, nor was the walking path. In this way, we could ensure that all participants walked along the same path and saw the same objects in exactly the same locations.

Map-pointing activity (Activity 2): The map-pointing activity required participants to complete a computer-based test, as shown in Fig. 2c. The test was designed as a desktop application in which participants were asked whether they had seen a particular object. If so, they were asked to mark its position on the corresponding 2D floor plan with a mouse click. The desktop test consisted of 15 questions. Ten questions were related to the ten objects presented to the participants during the object location memorization activity. Five questions were related to five objects that were not presented to the user during the object location memorization activity.

Each participant received instructions on how to use the desktop application during the map-pointing activity and a sample question prior to the computer-based test. Ideally, participants would only mark the locations of the objects that were shown to them during the object location memorization activity. For the five objects that were not shown to them during the object memorization activity, participants should indicate that they had not seen the objects. Participants were informed that they were not required to mark the exact positions of the memorized objects. Positions indicated within a circle with a radius of two meters around the positions recorded by the developed 3D positioning system were counted as correct responses. The desktop application recorded the time each participant took to complete the test.

Experimental procedures for the "with AR" condition: In the experimental session, the participant was asked to put on the AR headset and complete the object location memorization and map-pointing activities. At the beginning of the object location memorization activity, the researcher switched on the Unity-based application in the AR headset and helped the participant to put on the headset. While the participant walked along the three floors of the building with the AR headset (see Fig. 3b), the researcher followed the participant and wheeled a trolley containing the computing station (i.e., a laptop and a Wi-Fi router), as shown in Fig. 3c. The 3D object with an external representation of the objects and the path walked (see Fig. 3a) was updated as the participant followed the walking path. While walking, the participant was asked to hold their head slightly upward so that the AR markers attached to the ceilings along the walking path were in their field of view.

On the walking path, the participant saw different objects. Once they reached an object, the participant was instructed to stop in front of the object, first look at the marker on the ceiling near the object for one second, and then look at the object for another second. At that moment, the colored sphere with the label of the object was marked on the 2D virtual plan of the building, and was integrated to the real environment as a virtual 3D object in front of the user's eyes. As the participant walked along the path, they could see how the 3D object

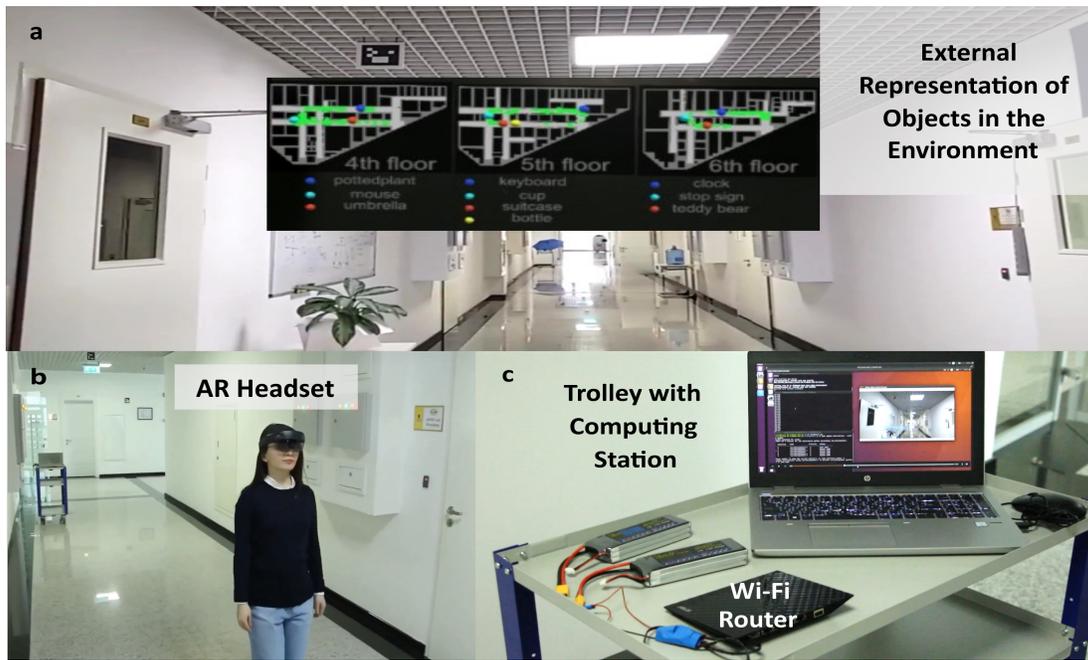


Fig. 3. Experimental procedure for the “with AR” condition (Activity 1: Object Location Memorization). (a) The external representation of the objects in the environment and the path walked. (b) The participant completes the walking tour of the corridors using the AR system. (c) The trolley with the computing station of the AR system wheeled by the researcher.

was updated in the AR environment. At the end of the object location memorization activity, the participant had the virtual 2D map of the building in front of their eyes, in which the objects and the walked path in the environment were drawn, as shown in Fig. 3a.

Once the object location memorization activity was completed, the participant was asked to remove the headset. The researcher switched off the Unity-based application, which created a digital representation of the objects in the environment and the user’s walked path. The participant was informed that the digital external representation of the objects in the environment, showing the object locations and the walked path, was saved in the memory of the AR headset, and that they could use the recorded information during the map-pointing activity.

After completing the object location memorization activity, the researcher and participant returned to the laboratory. The participant had a five-minute break and was asked to complete the Unweighted/Raw NASA-Task Load Index, also referred to as Raw TLX (RTLX) [47] and System Usability Scale (SUS) [48] questionnaires. These questionnaires were used to evaluate the cognitive load experienced by the participant during the object location memorization activity and the usability of the AR system, which was used to create the digital external representation of the objects placed in the environment.

While the participant was completing the questionnaires, the researcher switched on another Unity-based application on the AR headset that placed the 3D object with an external representation of objects at a fixed position in the real environment. When the user changed the position or turned their head in different directions, the virtual 3D object with an external representation of objects remained in the location where it was first initialized. For ease of use, the researcher switched

on the second application so that the virtual object was placed in front of the participant near the computer on which the computer-based test was running to complete the map-pointing activity (see Fig. 4b). The participant was asked to put on the AR headset and use the digital external representation of the objects in the environment to answer the test questions (see Fig. 4a). Later, the participant was asked to fill in the questionnaires describing their cognitive load after the map-pointing activity and evaluating the usability of the AR system. After a few minutes, the participant was asked to complete the same questionnaires, but this time describing their cognitive load after completing both activities. The participant was also asked to evaluate system usability for the entire experimental session, which consisted of performing two activities—that is, memorizing the location of objects within the building and later showing the location of the memorized objects on the map of the building.

Experimental procedures for the “without AR” condition: In the experimental session without AR, the participant did not use the AR system. The researcher accompanied the participant during the session to prevent them from getting lost. At the beginning of the object location memorization activity, the participant received the printed plans of the three floors of the building. The plans indicated the position of the elevators that could be used to move between the floors and the corridors that the participant had to walk along. The researcher indicated their position on the plan of the fourth floor when they started the experimental session. The participant was informed that they would first walk along the three floors of the building, memorize the locations of the objects they would see, and then return to the laboratory to complete a computer-based test. The printed plans of the three floors that were given to the participant were the same as those used

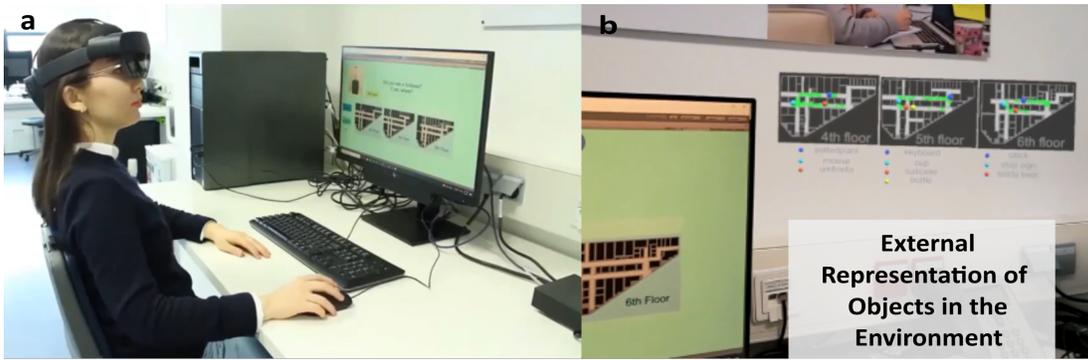


Fig. 4. Experimental procedure for the “with AR” condition (Activity 2: Map-Pointing). (a) A participant taking the computer-based test with the AR system. (b) The external representation of objects in the environment is integrated into the real environment as a virtual 3D object.

for the computer-based test. As the participant walked along the corridors of the three floors, they saw various objects on tables. Their task was to memorize the locations of the objects they saw using the floor plans provided. The participant was not given a precise strategy to memorize the objects and their corresponding locations. Ideally, the participant should locate their position on the floor plan and memorize the object nearby based on the floor plan.

After completing the object location memorization activity, the participant and researcher returned to the laboratory. The participant had a five-minute break and was asked to fill in the Unweighted/Raw NASA-TLX questionnaire to indicate the cognitive load experienced during the object location memorization activity. In the meantime, the researcher launched the test on the laboratory computer to continue with the map-pointing activity and asked the participant to answer the questions of the test. The participant could use the printed floor plans that they had received at the beginning of the object location memorization activity. Once the test was completed, the participant was asked to fill out the Unweighted/Raw NASA-TLX questionnaire to indicate the cognitive load they experienced after completing the map-pointing activity. A few minutes later, the participant was asked to complete the same questionnaire. This time, the participant was asked to rate the cognitive load they had felt after completing the two activities in the experimental session.

Participants: A total of 26 participants (12 females and 14 males) from the university community (students, researchers, and faculty members) were recruited to participate in the behavioral study. Their ages ranged between 20 and 39 years. We randomly divided participants into two gender-balanced groups: Group A (6 females and 6 males) and Group B (6 females and 8 males). The mean age (M) of participants in the two groups was 27 years with a standard deviation (SD) of 5 years ($M=27$ years, $SD=5$ in Group A and $M=27$ years, $SD=5$ in Group B). According to the Shapiro-Wilk test, the age of the participants in the two groups did not follow a normal distribution. The Mann-Whitney U test showed that the age difference between the two groups was not statistically significant ($U = 80.5$, $p = .44$). Group A completed the experimental task with and without AR assistance on two days (i.e., sessions) with a two-week break between the sessions. On Day 1, they completed the experimental task, which consisted

TABLE I
EXPERIMENTAL DATA COLLECTION PROCEDURE FROM THE PARTICIPANTS

	Day 1		Day 2	
	Activity 1 Object Location Memorization	Activity 2 Map Pointing	Activity 1 Object Location Memorization	Activity 2 Map Pointing
Group A	AR	AR	No AR	No AR
Group B	No AR	No AR	AR	AR

of two activities, with the AR system; on Day 2 (two weeks later), they completed the same experimental task without the system. Group B completed the same procedure in reverse order (i.e., on Day 1, they did not use the AR system, but, on Day 2, they did) (see Table I).

Ethical Approval. The Institutional Research Ethics Committee of Nazarbayev University approved the behavioral study with human participants. All study participants provided written informed consent.

C. Measures

1) *Subjective Assessment:* In both activities involved in the experimental task, we compared participants’ mental workload when they used the memory augmentation system and when they did not use it. To measure participants’ mental workload, we used the Unweighted/Raw NASA-Task Load Index (NASA-TLX), referred to as Raw TLX (RTLX) [47], as a subjective tool to assess mental workload. This index considers human cognitive workload across six dimensions, including mental demand, physical demand, temporal demand, performance, effort, and frustration. For these qualitative assessments, participants rated their experience after each task. In this way, we studied the impact of each activity on participants’ cognitive state across multiple dimensions. For each dimension, test scores ranged from 0 to 100 with a corresponding interpretation from “very low” to “very high”. The rating scale definitions used in NASA-TLX and further details can be found in [49]. RTLX results were analyzed using a dependent two-sample t-test (i.e., paired-samples t-test) ($p \leq 5\%$).

2) *Objective Assessment:* Along with the subjective method for determining cognitive workload, we also utilized objective methods for quantitative measurements. To this end, we used

performance evaluation metrics during the computer-based test in the map-pointing activity. The computer-based test in Fig. 2c was used to collect quantitative evaluation data. The desktop application recorded participants' responses to 15 questions and response times. The desktop application was designed on the Unity 3D 2021 game development platform and collected two sets of quantitative data: error rate and task completion time. If the user's response was more than two meters off the correct position, the system considered it an error. The error rate indicated the ratio of incorrect answers to all answers as a percentage. The task completion time indicated the total time it took each participant to complete the test. Mean error rates and response times were compared between the "with AR" and "without AR" conditions. The ratings were analyzed using a paired-samples t-test with a significance level of 5% for all comparisons.

3) *System Usability*: To evaluate the usability of the "ExoMem" system, we employed the SUS score, which is a composite measure of the overall usability of a system [48]. This is a subjective evaluation method of usability based on a post-task questionnaire completed after testing the system. The SUS score ranges from 0 to 100 and has corresponding interpretations along the usability and acceptability scales. The calculation of the score and further details can be found in [48].

IV. RESULTS

A. Subjective Assessment

The subjective assessment of mental workload for each activity performed by two groups of participants in two conditions with and without the AR system is shown in Fig. 5. Participants experienced much less mental demand and effort when using the AR system in both activities. Similarly, participants' experience of temporal demand and frustration decreased with the AR system in each activity. A paired-samples t-test revealed a significant effect of the AR system on cognitive demand [$t(50) = 10.19, p < .001$], temporal demand [$t(50) = 4.31, p < .001$], performance [$t(50) = 5.28, p < .001$], effort [$t(50) = 9.91, p < .001$], and frustration [$t(50) = 4.22, p < .001$]. A paired-samples t-test showed no significant effect of the AR system on physical demand [$t(50) = 1.47, p = .15$] in either activity.

B. Objective Assessment

Figure 6 shows participants' results on the map-pointing activity. We note that the error rate with "ExoMem" for both groups of participants was 3.85% (SD = 6.30), while the error rate without "ExoMem" was 28.97% (SD = 17.58). This shows that the error rate in the map-pointing activity was reduced 7.52 times when the AR system was used (see Fig. 6a). A significant difference in error rate was observed between the "with AR" and "without AR" conditions ($t(50) = 7.99, p < .001$). When using the AR system during the map-pointing activity, 46.7% of the errors resulted from incorrect operation of the system, and 53.3% of the errors were caused by participants incorrectly using the information presented to them by the system. The completion time of the computer-based test for

the map-pointing activity with "ExoMem" was 150.9 seconds (SD = 66.51), while without the AR system it was 206.1 seconds (SD = 105.52). In other words, task completion took 27% less time when the AR system was used (see Fig. 6b). A paired-samples t-test showed a significant difference between the "with AR" and "without AR" conditions in completion time ($t(50) = 2.36, p = .026$).

C. Effects of Gender on Performance in Objective Assessment

Studies of spatial tasks among children revealed that boys performed better on purely spatial tasks, whereas girls performed better on verbal tasks [50]. In the object location memorization task, boys and girls showed similar performance. However, female participants performed better than male participants in adulthood. In our work, we also examined whether the gender difference of our participants affected performance during the experimental task in our behavioral study. The error rate and completion times for male and female participants are illustrated in Fig. 7. An independent-samples t-test revealed no significant difference in total response time (including the "with AR" and "without AR" conditions) between genders ($t(50) = 0.104, p = .46$) and in total error rate between genders ($t(50) = 0.41, p = .34$).

D. Effects of the Experiment Order on Performance

To explore if the order of the conditions "with AR" or "without AR" could affect the results of the behavioral study, we randomly divided our participants into two groups. One group of participants (Group A) completed the experimental task with the AR system on the first day and without the AR system after two weeks. The other group (Group B), on the other hand, completed the experimental task without the AR system on the first day and with the AR system after two weeks. An independent-samples t-test showed that the order of conditions (i.e., "with AR" and "without AR") in our behavioral study (between Groups A and B) made no significant difference in error rate ($t(50) = -0.008, p = .49$) and completion time ($t(50) = -0.69, p = .25$).

E. System Usability

Figure 8 shows the system usability results reported by participants after they completed the object location memorization and map-pointing activities with the AR system. We found that the SUS scores calculated based on the responses averaged 83.5 (SD = 12.31) for Activity 1 (i.e., memorization of object locations), 90.2 (SD = 10.02) for Activity 2 (i.e., pointing the locations of the learned objects on the map), and 87.6 (SD = 10.98) for both activities overall in the two groups. According to Sauro et al. [51], a SUS score above 80.3 suggests that the system will be recommended to friends.

F. Correlation Analysis

We also explored the correlations between error rate, completion time, mental demand, effort, temporal demand, frustration, physical demand, performance, and SUS measures. The results of this correlation analysis are presented in Table II.

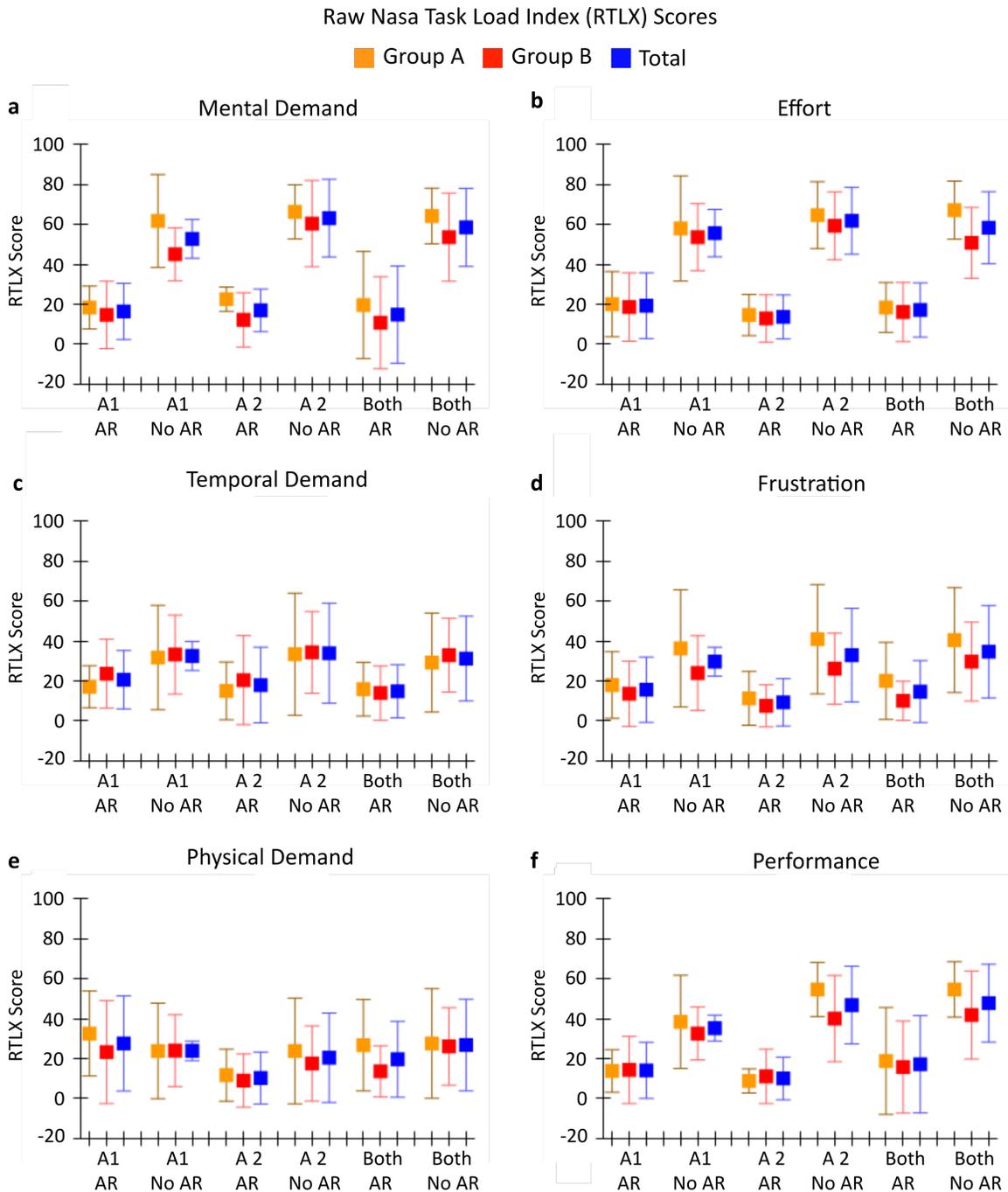


Fig. 5. **Average Raw NASA-TLX (RTLX) Scores.** Charts a-f show the mean and standard deviation values of average RTLX scores of participants in Group A ($n_A = 12$), Group B ($n_B = 14$), and overall ($n = 26$). Participants in Group A ($n_A = 12$) completed Activity 1 (A1) - Object Location Memorization and Activity 2 (A2) - Map-Pointing with AR system (AR) on Day 1 and without AR system (No AR) on Day 2. Participants in Group B ($n_B = 14$) completed Activity 1 (A1) - Object Location Memorization and Activity 2 (A2) - Map-Pointing without AR system (No AR) on Day 1 and with AR system (AR) on Day 2.

The correlation is considered weak when the absolute value of the correlation coefficient r is less than 0.3, moderate when r is between 0.3 and 0.5, and strong when r is higher than 0.5 [52]. In our case, the error rate was strongly and positively correlated with the time it took female participants to complete the map-pointing activity with the AR system. The completion time of the map-pointing activity completed with the AR system was also strongly and positively correlated with the temporal demand and performance of male partici-

pants. SUS measures of male participants were highly and negatively correlated with frustration. RTLX measures were generally correlated with each other. For instance, the mental demand and effort of both genders were highly and positively correlated in both conditions (i.e., “with AR” and “without AR”). Temporal demand had a strong positive correlation with the frustration and physical demand of female participants and with the physical demand of male participants in both conditions. Female participants’ frustration and physical de-

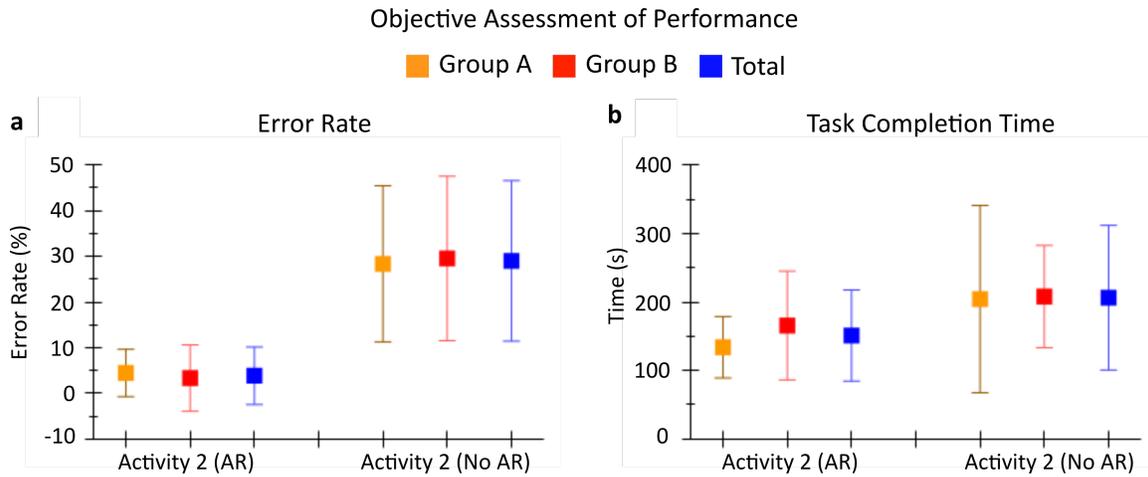


Fig. 6. Objective assessment of performance during map-pointing activity (Activity 2) with and without the AR system (AR and No AR). Charts a-b show the mean and standard deviation values of the average error rates and test completion times of participants in Group A ($n_A = 12$), Group B ($n_B = 14$), and overall ($n = 26$). (a) Average error rates (%). (b) Average task completion times (s).

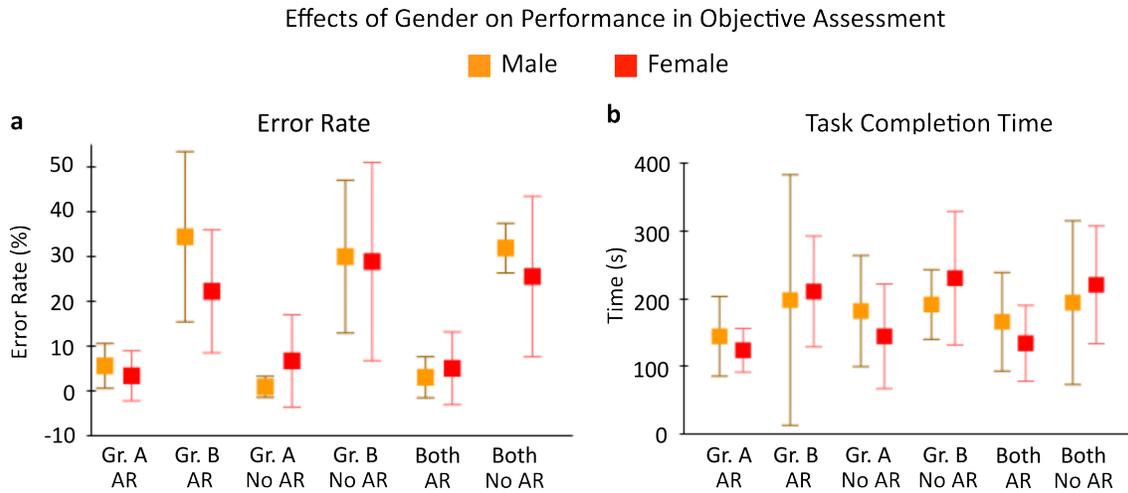


Fig. 7. Effects of gender on objective assessment of performance during map-pointing activity (Activity 2) with and without the AR system (AR and No AR). Charts a-b show the average error rates and test completion times of participants across genders in Group A ($n_A = 12$), Group B ($n_B = 14$), and overall ($n = 26$).

mand showed a strong positive correlation. Interestingly, male

participants that perceived higher mental demand were more likely to perform better.

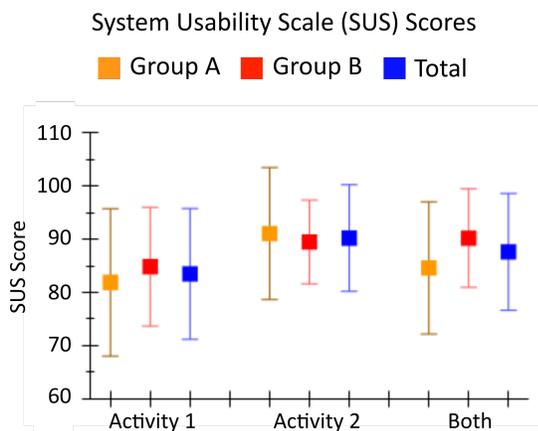


Fig. 8. System Usability. The chart shows the mean and standard deviation values of the average SUS scores rated by participants in Group A ($n_A = 12$), Group B ($n_B = 14$), and overall ($n=26$).

V. DISCUSSION

Over the last decades, cartographic data and mapping have moved into the realm of sensors (e.g., global navigation satellite system receivers and light detection and ranging (LiDAR) that support location-based applications to track the movement of people, vehicles, ships, and aircraft [35]. State-of-the-art AR head-mounted displays (HMDs) (e.g., Microsoft HoloLens 2) are equipped with a variety of sensing technologies (visible light cameras for head tracking, infrared cameras for eye tracking, a time-of-flight depth sensor, an inertial measurement unit (IMU), a video camera, a five-channel microphone array, built-in spatial sound speakers, and high-speed wireless communications) [53]. We sought to combine software and hardware features of the emerging technology (i.e., HoloLens 2 AR headset, CV, and AI) to develop an AI-supported AR system capable of sensing the environment and creating an external representation of objects in space to replace internal

TABLE II
CORRELATIONS BETWEEN OBJECTIVE AND SUBJECTIVE MEASURES

		Error rate	Completion Time	RTLX						SUS	
				Mental demand	Effort	Temporal demand	Frustration	Physical demand	Performance		
Error rate	AR	1	0.912	-0.081	0.286	-0.02	0.0381	0.008	-0.08	-0.292	
	No AR	1	-0.076	-0.156	-0.292	0.116	0.454	0.182	0.084	N/A	
Completion time	AR	-0.365	1	0.24	0.03	-0.23	0.153	0.187	-0.1	-0.203	
	No AR	0.121	1	-0.363	-0.123	0.111	-0.047	0.169	-0.02	N/A	
RTLX	Mental demand	AR	-0.05	0.34	1	0.529	0.93	0.628	0.671	0.006	0.14
		No AR	0.03	-0.15	1	0.683	0.301	0.277	0.407	0.154	N/A
	Effort	AR	-0.12	0.415	0.971	1	0.635	0.628	0.7	-0.329	-0.455
		No AR	-0.059	-0.228	0.914	1	0.156	0.079	0.274	0.482	N/A
	Temporal demand	AR	-0.171	0.59	0.911	0.919	1	0.643	0.601	-0.141	0.187
		No AR	-0.415	-0.324	0.204	0.220	1	0.708	0.711	-0.282	N/A
	Frustration	AR	0.027	0.127	0.884	0.885	0.786	1	0.75	0.01	-0.439
		No AR	-0.307	-0.138	0.197	0.157	0.011	1	0.698	0.112	N/A
	Physical demand	AR	-0.006	0.397	0.913	0.902	0.918	0.8137	1	-0.139	-0.479
		No AR	-0.321	-0.183	0.19	0.236	0.838	0.372	1	0.165	N/A
	Performance	AR	-0.254	0.735	0.648	0.649	0.831	0.350	0.698	1	0.248
		No AR	0.402	0.087	0.715	0.639	0.028	0.672	0.114	1	N/A
	SUS	AR	-0.011	0.027	-0.406	-0.372	-0.258	-0.527	-0.490	-0.028	1

Male and female participants' measures were highlighted with blue and red colors, respectively. Bold text values indicate strong correlations. SUS measures in the "No AR" columns are "N/A" in the table, because the SUS questionnaire was not taken when the experimental task, which consisted of two activities, was performed without the AR system.

mental representations. In designing the system, we relied on the internal mechanisms thanks to which people process visual and spatial information. We also presented the methodology of analysis of cognitive workload, performance, and usability in the experimental task in the behavioral study.

Our AR and AI-enhanced human memory augmentation system performed the following functions in the behavioral study: (1) acquiring FPV data from AR goggles for CV-based indoor localization and object detection, (2) establishing two-way communication between AR and the AI parts of the system through a wireless network. Our system integrated AR and AI technologies to sense the environment, exchange data over the wireless network, construct the external representation of objects in the indoor environment, and display the recorded data as a virtual 3D object in the AR environment. Based on the subjective evaluations of cognitive load and performance, we demonstrated that our system helped to reduce mental workload and improve performance in the experimental task, which involved two activities—namely, memorization of object locations and marking the locations of the learned objects on the map. Statistical analysis of the records of error rate and time to complete the map-pointing activity in the two conditions (with and without AR) revealed no significant differences between the genders. We also analyzed the correlations between the reported variables (i.e.,

error rate, completion time, mental demand, effort, temporal demand, frustration, physical demand, performance, and usability scores). Interestingly, error rate showed a strong and positive correlation with completion time for female participants. In contrast, for male participants, there was a strong and positive correlation between completion time and temporal demand and performance during the map-pointing activity performed with the AR system.

Compared to a controlled research environment, realistic indoor environments posed technical challenges for our system. Firstly, excessive natural light through windows negatively affected the operation of the AR headset, preventing it from correctly sensing the environment [54]. Specifically, once perception problems arose, the performance of the AI part of the system also degraded. This led to the failure of user localization and object detection. Secondly, the user localization method also had its inherent technical limitations. In particular, our implementation depended on ArUco fiducial markers [44], which had to be placed in the indoor environment and whose coordinates were recorded in advance. Moreover, such user localization method was subject to inaccuracies, especially along the vertical z-axis, demonstrating that the calibration process of the camera device is vital for localization in multi-story buildings. Finally, the object detection method used also had some drawbacks. While large- and medium-sized

objects could be recognized from a distance of two meters, the detection of smaller objects (e.g., the computer mouse in our experiments) required a distance up to three times shorter. Thus, the system became less comfortable for users when more small objects were present in the experimental scenario.

In addition, there were limitations in our approach related to the experimental procedure. For example, the wearable and AI modules of the system were switched on and off by the instructor during the two activities in the experimental task. In the object location memorization activity, the instructor followed the users along the path and checked whether the localization markers and objects were correctly recognized. In the map-pointing activity, the instructor switched on and off the application that loads and displays the virtual 3D object with the external representation of the object in the environment using an AR headset. In the real world, users themselves would need to learn how to turn on and off the wearable and AI modules of the system. Such a learning process could increase the user's cognitive load and affect the overall usability of the system. In future experiments, participants should use the system autonomously to obtain a more realistic cognitive load and performance evaluation.

According to Clarke [27], the ideas that are considered innovative and fresh today tend to inspire and promote the creation of new approaches, methods, and technologies in the future. Further integration of digital interactivity into our daily lives requires a detailed analysis of the perceptual, cognitive, cultural, and practical aspects that influence the human experience with interactivity and visualization [30], [31]. Roth et al. [31] presented emerging interdisciplinary recommendations for future research related to user studies in cartography with a focus on interactive maps and visualization. According to the authors, usability of such interactive visualization systems requires that users and designers meet at the interface to provide a positive experience. Interactivity requires that users no longer remain passive in the visual representation of spatial information. As Muehlenhaus et al. [32] noted, map interactivity allows users to create a representation that best suits their goals and context of use.

In general, spatial perception and memorization tasks require mental and cognitive demands. Although our understanding of human cognition and memory is improving and is supported by technological breakthroughs in many areas, such as computing power, computer vision, and artificial intelligence, assistive solutions to augment human memory are still in the prototype stage and mainly focus on training the memory or supporting the memorization process [25]. New solutions to replace internal representation with the external technology generated are yet to be developed. The results of our work highlight the potential of AR and AI technologies in developing a new generation of intelligent technological systems capable of supporting human cognition by reducing the need to rely only on the internal representation of spatial information. For future researchers who wish to further develop the presented approach to human memory augmentation, we suggest conducting a behavioral study involving more individuals from different demographic groups and supplementing the study with structured interviews. Structured interviews help

participants share their ideas and experiences after using the system, as well as their overview and understanding of the concepts being explored.

VI. CONCLUSION

This paper presents a new approach to augmenting human memory using AR and AI technologies. Our AR system constructs the external augmented representation of objects located in the indoor environment based on the user's experience. The system is comprised of wearable AR and AI computing modules. The system was validated in a behavioral study with 26 human participants who completed an experimental task consisting of two activities (i.e., memorization of object locations in the environment and showing the locations of the learned objects on the map) under two conditions (i.e., with and without our AR system). Evaluation of cognitive load showed that participants experienced lower cognitive load when using the system. During the map-pointing activity, participants made 7.52 times fewer errors on the post-memorization computer-based test when the system was used. The usability evaluation of the system yielded a SUS score of over 80% among the participants. The statistical analysis of the error rate and the completion time recorded during map-pointing activity revealed no significant difference between the genders.

ACKNOWLEDGMENTS

We thank Askat Kuzdeuov for developing the ArUco marker-based camera localization framework for three floors of the C4 research building at Nazarbayev University.

DATA AVAILABILITY

The data collected from participants in our experimental studies are available at <https://drive.google.com/drive/folders/1d0y73MaLYDGnXHFyNfsMOMxcvIwlrBDm>.

CODE AVAILABILITY

All the codes developed for the memory augmentation system and experiments are available at <https://github.com/IS2AI/ExoMem-AR-Memory>.

REFERENCES

- [1] Y.-H. Lee, T. Zhan, and S.-T. Wu, "Prospects and challenges in augmented reality displays," *Virtual Reality & Intelligent Hardware*, vol. 1, no. 1, pp. 10–20, 2019.
- [2] M. Tönnis, D. A. Plecher, and G. Klinker, "Representing information-classifying the augmented reality presentation space," *Computers & Graphics*, vol. 37, no. 8, pp. 997–1011, 2013.
- [3] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.
- [4] T. Zhan, K. Yin, J. Xiong, Z. He, and S.-T. Wu, "Augmented reality and virtual reality displays: perspectives and challenges," *Iscience*, vol. 23, no. 8, p. 101397, 2020.
- [5] J. Keil, A. Korte, A. Ratmer, D. Edler, and F. Dickmann, "Augmented reality (AR) and spatial cognition: Effects of holographic grids on distance estimation and location memory in a 3D indoor scenario," *PGF—Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 88, no. 2, pp. 165–172, 2020.

- [6] C. Carbonell Carrera and L. A. Bermejo Asensio, "Augmented reality as a digital teaching environment to develop spatial thinking," *Cartography and Geographic Information Science*, vol. 44, no. 3, pp. 259–270, 2017.
- [7] H. Chen, L. Hou, G. K. Zhang, and S. Moon, "Development of BIM, IoT and AR/VR technologies for fire safety and upskilling," *Automation in Construction*, vol. 125, p. 103631, 2021.
- [8] R. M. Kitchin, "Cognitive maps: What are they and why study them?" *Journal of Environmental Psychology*, vol. 14, no. 1, pp. 1–19, 1994.
- [9] N. Burgess, "Spatial cognition and the brain," *Annals of the New York Academy of Sciences*, vol. 1124, no. 1, pp. 77–97, 2008.
- [10] A. D. Baddeley, *Human memory: Theory and practice*. Psychology Press, 1997.
- [11] A. Conway, C. Jarrold, M. Kane, A. Miyake, and J. Towse, *Variation in working memory*. Oxford University Press, 2007.
- [12] A. Baddeley, "Working memory: Looking back and looking forward," *Nature Reviews Neuroscience*, vol. 4, no. 10, pp. 829–839, 2003.
- [13] E. E. Smith and J. Jonides, "Working memory: A view from neuroimaging," *Cognitive Psychology*, vol. 33, no. 1, pp. 5–42, 1997.
- [14] A. Postma, R. P. Kessels, and M. van Asselen, "How the brain remembers and forgets where things are: The neurocognition of object–location memory," *Neuroscience & Biobehavioral Reviews*, vol. 32, no. 8, pp. 1339–1345, 2008.
- [15] J. K. O'regan, "Solving the "real" mysteries of visual perception: the world as an outside memory," *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, vol. 46, no. 3, p. 461, 1992.
- [16] N. Cowan, "The magical mystery four: How is working memory capacity limited, and why?" *Current Directions in Psychological Science*, vol. 19, no. 1, pp. 51–57, 2010.
- [17] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
- [18] J. Sweller, "Cognitive load theory: Recent theoretical advances," 2010.
- [19] J. J. Van Merriënboer and J. Sweller, "Cognitive load theory and complex learning: Recent developments and future directions," *Educational Psychology Review*, vol. 17, no. 2, pp. 147–177, 2005.
- [20] H. Atıcı-Ulusu, Y. D. İkiz, O. Taskapilioglu, and T. Gunduz, "Effects of augmented reality glasses on the cognitive load of assembly operators in the automotive industry," *International Journal of Computer Integrated Manufacturing*, vol. 34, no. 5, pp. 487–499, 2021.
- [21] M. Thees, S. Kapp, M. P. Strzys, F. Beil, P. Lukowicz, and J. Kuhn, "Effects of augmented reality on learning and cognitive load in university physics laboratory courses," *Computers in Human Behavior*, vol. 108, p. 106316, 2020.
- [22] M. Aebbersold, T. Voepel-Lewis, L. Cherara, M. Weber, C. Khouri, R. Levine, and A. R. Tait, "Interactive anatomy-augmented virtual simulation training," *Clinical Simulation in Nursing*, vol. 15, pp. 34–41, 2018.
- [23] M.-C. Juan, M. Mendez-Lopez, E. Perez-Hernandez, and S. Albiol-Perez, "Augmented reality for the assessment of children's spatial memory in real settings," *PLoS One*, vol. 9, no. 12, p. e113751, 2014.
- [24] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood, "Sensecam: A retrospective memory aid," in *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County, CA, USA, September 17-21, 2006 Proceedings 8*. Springer, 2006, pp. 177–193.
- [25] M. Harvey, M. Langheinrich, and G. Ward, "Remembering through lifelogging: A survey of human memory augmentation," *Pervasive and Mobile Computing*, vol. 27, pp. 14–26, 2016.
- [26] A. L. Griffin, A. C. Robinson, and R. E. Roth, "Envisioning the future of cartographic research," *International Journal of Cartography*, vol. 3, pp. 1–8, 2017.
- [27] K. C. Clarke, J. M. Johnson, and T. Trainor, "Contemporary American cartographic research: A review and prospective," *Cartography and Geographic Information Science*, vol. 46, no. 3, pp. 196–209, 2019.
- [28] M. Yuan and K. S. Hornsby, *Computation and visualization for understanding dynamics in geographic domains: a research agenda*. CRC Press, 2007.
- [29] R. A. White, A. Coltekin, and R. R. Hoffman, *Remote sensing and cognition: Human factors in image interpretation*. CRC Press, 2018.
- [30] A. Çöltekin, I. Lokka, and M. Zahner, "On the usability and usefulness of 3d (geo)visualizations - a focus on virtual reality environments," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B2, pp. 387–392, 2016.
- [31] R. E. Roth, A. Çöltekin, L. Delazari, H. F. Filho, A. Griffin, A. Hall, J. Korpi, I. Lokka, A. Mendonça, K. Ooms *et al.*, "User studies in cartography: opportunities for empirical research on interactive maps and visualizations," *International Journal of Cartography*, vol. 3, no. sup1, pp. 61–89, 2017.
- [32] I. Muehlenhaus, *Web cartography: map design for interactive and mobile devices*. CRC Press, 2013.
- [33] J. Bobrich and S. Otto, "Augmented maps," *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 34, no. 4, pp. 502–505, 2002.
- [34] S. K. Pal, A. Pramanik, J. Maiti, and P. Mitra, "Deep learning in multi-object detection and tracking: State of the art," *Applied Intelligence*, vol. 51, no. 9, pp. 6400–6429, 2021.
- [35] J. M. Dow, R. E. Neilan, and C. Rizos, "The international GNSS service in a changing landscape of global navigation satellite systems," *Journal of Geodesy*, vol. 83, no. 3, pp. 191–198, 2009.
- [36] R. F. Brena, J. P. García-Vázquez, C. E. Galván-Tejada, D. Muñoz-Rodríguez, C. Vargas-Rosales, and J. Fangmeyer, "Evolution of indoor positioning technologies: A survey," *Journal of Sensors*, vol. 2017, pp. 1–22, 2017.
- [37] J. Polvi, T. Taketomi, G. Yamamoto, A. Dey, C. Sandor, and H. Kato, "SlidAR: A 3D positioning method for SLAM-based handheld augmented reality," *Computers & Graphics*, vol. 55, pp. 33–43, 2016.
- [38] U. Rehman and S. Cao, "Augmented-reality-based indoor navigation: A comparative analysis of handheld devices versus google glass," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 1, pp. 140–151, 2016.
- [39] D. Damen, T. Leelasawassuk, and W. Mayol-Cuevas, "You-Do, I-Learn: Egocentric unsupervised discovery of objects and their modes of interaction towards video-based guidance," *Computer Vision and Image Understanding*, vol. 149, pp. 98–112, 2016.
- [40] A.-M. Calle-Bustos, M. Méndez-López, J. Juan, P. Dias, F. Abad, and M. Carmen Juan, "Visual vs auditory augmented reality for indoor guidance," in *Proc. of the International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISI-GRAPP)*, vol. 1: GRAPP, 2021, pp. 85–95.
- [41] F. Muñoz-Montoya, M.-C. Juan, M. Mendez-Lopez, R. Molla, F. Abad, and C. Fidalgo, "SLAM-based augmented reality for the assessment of short-term spatial memory: A comparative study of visual versus tactile stimuli," *PLoS ONE*, vol. 16, no. 2, p. e0245976, 2021.
- [42] N. R. Caluya, A. Plopski, J. F. Ty, C. Sandor, T. Taketomi, and H. Kato, "Transferability of spatial maps: Augmented versus virtual reality training," in *Proc. of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 387–393.
- [43] Y. Khassanov, M. Nurpeiissov, A. Sarkytbayev, A. Kuzdeuov, and H. A. Varol, "Finer-level sequential WiFi-based indoor localization," in *Proc. of the IEEE/SICE International Symposium on System Integration (SII)*, 2021, pp. 163–169.
- [44] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [45] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [46] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. of the European Conference on Computer Vision*. Springer, 2014, pp. 740–755.
- [47] S. G. Hart, "Nasa-task load index (NASA-TLX): 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 9, pp. 904–908, 2006.
- [48] J. Brooke, "SUS: A quick and dirty usability scale," *Usability Evaluation in Industry*, vol. 189, no. 3, 1996.
- [49] "NASA TLX: Paper/pencil version," 2019. [Online]. Available: <https://humansystems.arc.nasa.gov/groups/tlx/tlxpaperpencil.php>
- [50] P. A. Lowe, J. W. Mayfield, and C. R. Reynolds, "Gender differences in memory test performance among children and adolescents," *Archives of Clinical Neuropsychology*, vol. 18, no. 8, pp. 865–878, 2003.
- [51] J. Sauro and J. R. Lewis, *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.
- [52] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 2013.
- [53] "Microsoft HoloLens," 2021. [Online]. Available: <https://www.microsoft.com/en-us/hololens>
- [54] A. Erickson, K. Kim, G. Bruder, and G. F. Welch, "Exploring the limitations of environment lighting on optical see-through head-mounted displays," in *Proc. of the Symposium on Spatial User Interaction*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–8.