Journal of Universal Computer Science, vol. 24, no. 2 (2018), 108-128
submitted: 29/9/17, accepted: 3/11/17, appeared: 28/2/18

From AR to Expertise: A User Study of an Augmented Reality Training to Support Expertise Development

Bibeg Hang Limbu
(Welten Institute – Research Centre of Learning, Teaching and Technology
Open University of the Netherlands, Heerlen, Netherlands
Bibeg.limbu@ou.nl)

Halszka Jarodzka
(Welten Institute – Research Centre of Learning, Teaching and Technology
Open University of the Netherlands, Heerlen, Netherlands
Lund University, Sweden
Halzka.Jarodzka@ou.nl)

Roland Klemke
(Welten Institute – Research Centre of Learning, Teaching and Technology
Open University of the Netherlands, Heerlen, Netherlands
Roland.Klemke@ou.nl)

Fridolin Wild
(Performance Augmentation Lab, Department of Computing and Communication Technologies
Oxford Brookes University, Oxford, United Kingdom
Wild@brookes.ac.uk)

Marcus Specht
(Welten Institute – Research Centre of Learning, Teaching and Technology
Open University of the Netherlands, Heerlen, Netherlands
Marcus.specht@ou.nl)

Abstract: Augmented reality and sensor technologies have been analysed extensively in several domains including education and training. Although, varieties of use cases and applications exist, these studies were conducted in controlled laboratory environments. This paper reports on the first user study of augmented reality prototype developed to support students to learn from trainers in professional domains using augmented reality and sensors. The prototype records the performance of trainers in the first phase to support students by making it available during practice in the second phase. The performance data is made available to both the students and trainers in the third phase for reflection. A total of 142 participants which included trainers and students from three professional domains, namely 1) aircraft maintenance 2) medical imaging and 3) astronaut training, evaluated the prototype. The trainers used the prototype to record their performance while the students used the prototype to learn from the recorded performance. Participants from the three professional domains evaluated the usability of the prototype by means of a questionnaire. Randomly selected participants were also interviewed to collect their opinions and suggestion for further usability improvement. Furthermore, they also evaluated the implementation of the instructional design methods, which were identified prior in a literature review, with a brief questionnaire. The questionnaire was designed to measure the acceptance of the implementation of instructional design methods and to evaluate its adherence to the authors definition. The results of this study show that the usability of the prototype is below expected standard acceptable level. The results of the questionnaire on the implementation of the instructional design methods varied show above average acceptance levels by both the trainers and the students in the three professional domains. To conclude, the prototype shows potential to be used in different domains to support expertise development.

Keywords: augmented reality; expertise; training; instructional design
Categories: L.2.1, L.3.6, L.7.0

1 Introduction

Sensors and augmented reality (AR) technologies have been developing fast with several plateaus of maturity being observed, such as with the release of Microsoft Hololens, over the past years. However, sensors and AR suffer from various constraints that obstruct their optimal implementation in industrial and educational contexts. Instructional design issues such as the distribution and flow of information between the physical and the virtual environment and between different devices is still obscure [Wu 2013]. As a consequence, designing a training environment based on sensors and AR for facilitating expertise development is challenging [Drljevic 2017]. The complexity of interacting with large amounts of information and various devices at the same time, while performing a complex task can be overwhelming for the students. Designers of AR training environments need to realize the limitations to design the best possible training environment. In this regard, we adopt a design-based research approach [diSessa 2004] which allows the end users to be a part of the design process ensuring that the final product meets the user requirements and needs. This article presents the first user study performed with our prototype designed for supporting expertise development in students in professional domains.

Attaining expertise is a difficult endeavour with claims that it may take up to 10 years to become an expert [Gladwell 2008]. [Ericsson 2007] have emphasized the importance of experts as trainers for supporting expertise development. While trainers are imperative for supporting expertise development in an students, learning from them is difficult [Hinds 1999; Patterson 2010]. Moreover, access to trainers is limited for students which impedes their development even further. Various efforts that can be potentially translated to address these problems have been made in the last years. [Jarodzka 2017] have presented eye tracking sensors as tools for supporting instructional design and expertise development in various domains such as chess and medicine. Similarly, posture sensors have been used for training public speaking skills by [Schneider 2017]. In addition, numerous similar studies have presented the potential of sensors and AR based training systems for expertise development [Olwal 2008].

Sensors and AR have the capability to unobtrusively measure physical properties. The prototype used in this study utilizes the potential of sensors and AR to record trainer's performance. This recorded performance is used to train students with the help of sensors and AR by making it available to the students when needed. By doing so with the help of Sensors and AR, the prototype supports technology enhanced training in authentic context which facilitates expertise development in students [Carey 2014]. In addition, sensors and AR also have the potential to provide personalized feedback and guidance in real time [Bacca 2014]. These potentials of sensors and AR are crucial aspects for supporting expertise development in students [Ericsson 2006].

2 Sensors and AR for learning from trainers: the training methodology

To support expertise development, the training methodology of the prototype utilizes the valuable experience and knowledge that the trainer possesses. It intends to make the experience and knowledge of the trainer accessible and available to students. This methodology consists of three major phases: recording trainer's performance, re-enacting trainer's performance by the student, and reflection. The recording phase ensures that the trainer records all the relevant information needed for a student to perform the task. The re-enactment enables the student to learn from the recorded performance while the reflection phase allows the trainer and the student to reflect on the performance by observation or/and from the data collected.

The training methodology is supposed to be applicable across different domains. To meet this criterion, we identified diverse attributes, such as speed and accuracy, important in various domains. Initially, a literature review and interview of the trainers in three professional domains, namely 1) aircraft maintenance 2) medical imaging and 3) astronaut training, were conducted to identify the attributes [Limbu 2017]. IDMs used by the author in the studies reviewed, were extracted to support the training of each attribute (see Table 1). The commonly identified IDMs across all domains, as identified by the trainers in the domains, were implemented in the prototype used in this study. Table 1 provides the description of each of these IDMs along with how the trainer data was created which is outlined in the recording column. Similarly, the replay column describes how the trainer data was used for training.

Each IDM is mined from the literature review and the implementation is defined by the authors. IDMs defined in the context of this study, utilize recorded trainer's performance for training certain attributes of a skill with the help of sensors and AR. In contrast, the prototype described in this paper implements a pool of IDMs. Combining a pool of IDMs allows many inter-related aspects of a complex task to be trained. The prototype implements various IDMs together into a system which brings forth new challenges such as the usability of the system or even the assurance that each IDM implementation accomplishes its purpose. In addition, IDMs are abstract definitions and implementation methods can vary across platforms. Thus, we implement a collection of IDMs to test with end users in authentic settings and report the results of the first user study in this article.

In this study, we explore how the end users, that is the trainers and the students, from three professional domains perceive our prototype. To do so, we evaluate the prototypical implementation of the IDMs in three professional domains. It should be noted that the prototype is still in an early phase of development and thus measurements of effectiveness in training are not expected to be optimal. Therefore, in this study we test the following hypotheses:

1. System usability scale [Brooke 2013] is at an acceptable level of 70 from both the trainers' and the students' perspective.
2. The prototypical implementation of each IDM will meet the authors' defined purpose of training certain attribute from the trainers' and the students' perspective.

3. The prototypical implementation of collection of IDMs will be equally accepted in all domains, both, from the trainers' and the students' perspective.

Instructional Design Methods	Description	Methods and means used for recording	Methods and means used for replay
Highlight Object of Interest	Highlight physical objects in the visual area indicating the students that the trainer found that object of interest	Interview with trainers to determine the object of interest	Hololens highlights the location of the object by using a virtual interface
Directed Focus	Visual pointer for trainer determined relevant objects outside the visual area	Interview with trainers to determine objects of interest and observation from demonstration	Visual direction indicated by an arrow to direct attention
Point of View Video	Provides unique student/trainer point of view video which may not be available in a third person perspective	Head mounted camera in Hololens are used to let the trainer record videos Controls to initiate and stop recording	Video projected by Hololens in the relevant physical location
Think aloud	Audio recordings the explanations and mental process (think aloud protocol) of the trainer during the task execution	Built in microphones in Hololens used to record trainer's explanations Noise cancellation	Built in microphone for voice commands Built in Hololens speaker plays the recorded audio in relevant time and space
Cues & Clues	Cues and clues are pivots that trigger solution search. It can be in form of image or audio. It should represent the solution with a single annotation	Picture, audio or text are used to provide hints to the student Materials are identified during interview with trainers or during the demonstration of the task	The chosen media contents are displayed using Hololens in the relevant time and location

Instructional Design Methods	Description	Methods and means used for recording	Methods and means used for replay
Annotations	Allow a physical object to be annotated by the trainer during task execution. (Similar to sticky notes, but with more modes of information)	Gesture based interactions to tag media to a physical object. Performed by the trainer during demonstration	Camera to find the physical location of the annotations AR display unobtrusively relays the information
Object Enrichment	Provide domain related information about the physical artefact which are crucial to the performance of the task from a trainer's point of view	Interview with trainers determined relevant pieces of domain information apart from procedural information	Vuforia image recognition used to display such information in precise physical location
Contextual Information	Provide information about the process that is frequently changing but is important for performance.	Procedural information of the task is determined from the interview with trainer	Voice command based intractable checklist of steps to be performed is provided
3D Models and Animation	3d models and animations assist in easy interpretation of complex models and phenomena which require high spatial processing ability.	Modelling 3d object and creating 3d animation Interview with trainers to determine required models	Hololens display the 3d model which are moveable so that it is not obtrusive
Ghost track	Enables visualizing the recorded movement of the person's whole body	Trainers body movement is recorded while demonstrating the task	Hololens enabled visualization of recording through a holographic body enacting the recording.

Table 1: Implemented IDMs in the tested prototype

3 Method

3.1 Use cases and application domains

The prototype was tested in three different professional domains, namely 1) aircraft maintenance 2) medical imaging and 3) astronaut training. The aircraft maintenance training task consisted of ten steps of pre-flight inspection on an aircraft. Pre-flight inspection is used to determine if the aircraft is in an airworthy condition. Conducting a pre-flight inspection requires a lot of paperwork and reference information to be gathered and studied before proceeding to the aircraft to conduct the inspection. The ten steps in the pre-flight inspection require the participant to move along the aircraft cabin inspecting critical points for any hazards (see Figure 1).

Figure 1: Participant performing pre-flight inspection inside the cabin

The medical imaging focused on the training of radiologist students to perform an echographic examination by using an ultrasound machine. Unlike the aircraft maintenance, the participants are bound to a fixed location and the ultrasound machine which provides all the diagnostic information (see Figure 2). The participants require operating knowledge of the machine, including the process of examining the patient with the ultrasound, and the perceptual abilities to recognize any deformities in the images produced by the machine.

Figure 2: Participant examining a patient using ultrasound machine.

Similarly, the astronaut training was conducted with the installation of the temporary stowage rack in the automated transfer vehicle. The procedure was performed on a mock up vehicle where the participants were required to install the rack using the proprietary installation units provided (see Figure 3). The participants needed to know the location and the application procedure of the installation units which support the racks. Repetitive training sequences are needed to prepare the astronauts for all the activities and procedures required in space missions. These types of training practices accumulate to a large amount which takes significant proportion of training time.

3.2 Participants

The aircraft maintenance session in Lufttransport, Norway consisted of 31 students and 24 trainers. The students group comprised of student volunteers from bachelor programs of 'safety and environment', 'nautical sciences', and 'aviation' from the department of engineering & safety at The Arctic University of Norway. The trainers comprised of maintenance apprentices, skilled workers (mechanics) and technicians working in at Lufttransport. Similarly, 17 trainers and 22 students were involved in the astronaut training sessions in Altec, Italy. The trainers were Altec and Thales Alenia Space employees while the students were from the master in space exploration and development systems courses. During the medical imaging sessions in Ebit, Italy, 9 trainers varying from teachers to Medical doctors and 39 students from the faculty of medicine and ICT engineering participated in the session. Over all, in all three professional domains, there were 39 females and 103 males with most students age falling between 18-24 while most trainers age fell between 25-34.

Figure 3: Installation of the rack by the participant for astronaut training domain

3.3 Materials: AR and sensor prototype

The prototype aims to utilize the valuable experience and knowledge that trainers possess and make it accessible and available to all students. To achieve this, the prototype consists of two major components: the *recorder* to capture trainer's performance and the *player* for supporting training of students (see Figure 4). The recorder ensures that the trainer records all the relevant information needed to support the training of students. The player enables students to learn from the captured performance.

The prototype has been implemented for the Hololens which is an AR glass from Microsoft™. The recorder component uses various sensors depending on the requirement of the domain to record trainer's performance. It also allows trainers to create learning materials in authentic contexts as shown in Figure 6. In Figure 6 the trainer can annotate the physical object with a virtual information. The trainer can interact with the recorder using gestures to annotate a physical location with various types of data. For example, the recorder allows trainers to record audio, take pictures and place 3D models at various physical locations. At the end of each recording, the data from the recorder is stored and fed into the player.

The player component on other the hand is catered for the students. The students receive step by step auditory and visual instructions which guides and supports them through the task. The contents created by the trainers such as notes are projected on their relevant physical locations and time based on the data from the recorder as shown in Figure 7. students can also interact with the player using voice commands and gestures. The player allows students to navigate between the steps in the procedure using keyword based voice recognition. Both the recorder and the player are in the early stages of development. However, this study is more concerned with the usability of the system and the adherence of the implementation of IDMs according to the definition provided by the authors.

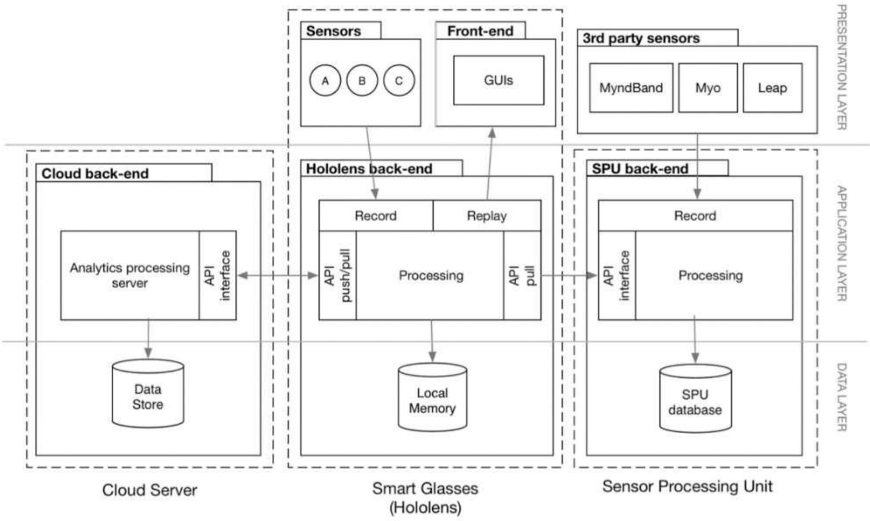

Figure 4: Architecture of the prototype

The prototype is developed in a three-layered architecture:

Presentation layer: the front-end and top-most level of the application, which consists of the graphical user interfaces (GUIs) and the sensor components to interact with the user and the external environment.

Service layer: the back-end and middle layer which coordinates the recorder and player clients, the data collection and analysis and the communication and transfer of these data across the platforms.

Data layer: the bottom layer where the information is stored such that it can be retrieved, processed, and re-presented to the user.

In addition to the three layers, the architecture combines three main computing units:

Hololens: The main wearable device through which the trainer can record his/her performance such that the learner can access it later. The Hololens will run both the two main applications of the prototype: the recorder and the player.

Sensor Processing Unit (SPU): The portable computer device works as hub for the third-party sensors that are not embedded in the smart glasses but are necessary for capturing performance. The SPU is responsible only for the receiving and recording of all the third-party sensors. In addition, it also offers the necessary API interfaces to allow the Hololens to retrieve and store sensor data.

Cloud Server: The cloud-based server is the place in which the recorded performances are saved and processed for later re-enactment. The cloud-based solution allows for a scalable and distributed data storing over a nearly infinite number of computer nodes, as well as the availability of the data to all the connected and authorised devices.

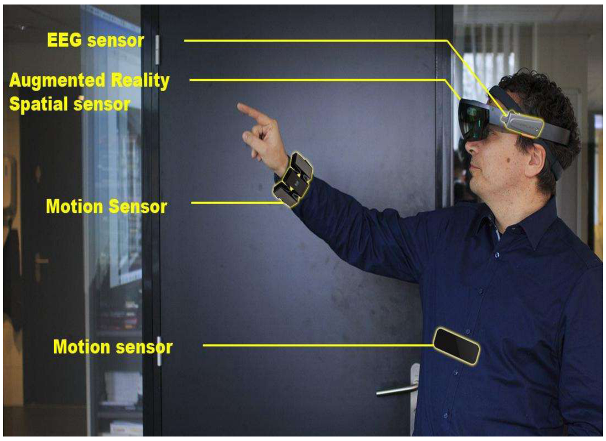

Figure 5: Physical setup of the prototype

Figure 6: Trainer's vision from the recorder for manually creating learning content

3.4 Procedure

The participants were scheduled to arrive in a group of 2-4 participants per hour. They were initially introduced to the project and asked to sign a consent form. They were requested to fill in a demographic questionnaire prior to the evaluation study. Trainers were exposed to both the recorder and the player while the students were only exposed to the player. However, the students were informed of the scenario during the briefings that the content they saw was created by the trainers during the recording phase. Both the students and the trainers were familiarized with the user interactions on the Hololens by means of inbuilt gesture training in Hololens. After they completed the gesture training, the trainers were required to use the recorder under supervision to ensure that they were familiarized with the recorder. This session was followed by a briefing which involved only the trainers, on what they were expected to do. Finally, the trainers were asked to demonstrate the assigned tasks from their domain. A printed list of steps was also provided to the trainers for reference. After the recording, the trainers were briefed on the player aspects of the prototype. The students were not required to use the recorder. Instead, the students immediately exposed to the player after the gesture training. Finally, participants completed the questionnaire containing questions about the IDMs and system usability which was measured by Standard Usability Scale questionnaire [Brooke 2013].

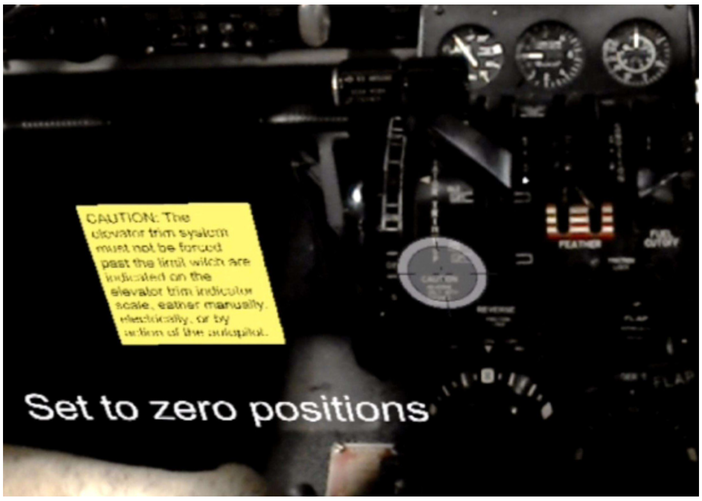

Figure 7: Student's view of player captured from the Hololens

The IDM questionnaire evaluated the IDMs to measure their adherence to the intended definition of the IDM by the author. In the IDM questionnaire, students and trainers were asked to rate the statements on a Likert scale of 1-7 based on their experience after using the prototype. The participants rated these statements between completely agree and completely disagree based on their experience. The statements were derived from the description of each IDMs. Each statement represented an ideal experience of the implementation of the corresponding IDM. Similarly, to measure system usability SUS was used. SUS is an industry standard tool for measuring the system usability in a quick manner. The SUS scores calculated from individual questionnaires represent the system usability. SUS yields a single number between 0 to 100 [Brooke 2013] representing a composite measure of the overall usability of the system being studied. Scores for individual items are not meaningful on their own. The acceptable SUS score is about 70 [Bangor 2009; Brooke 2013].

Sessions for each professional domain were held at their corresponding sites, with a week dedicated to each of them for preparation and execution. During the first session which is the aircraft maintenance, general technical issues and bugs in the prototype which affected the study directly were identified. These issues were resolved in the following sessions in case of astronaut training and medical imaging.

4 Results

System Usability Scale (SUS)

	Aircraft maintenance	Astronaut training	Medical imaging
Trainers	59.1(1.46)	69.2(1.06)	66.4(1.65)
Students	66.7(1.01)	67.5(1.83)	68(1.04)

Table 2: SUS scores in all the sessions

The average SUS score for aircraft maintenance for trainers (59.1) is below 70 which indicates that the recorder's usability is not on an acceptable level yet. There is a noticeable improvement in the SUS scores of the recorder between trainers from aircraft maintenance session and the other two sessions (Table 2). Amendments made to the recorder after the first session i.e. aircraft maintenance may have resulted in the improved SUS scores for the recorder in astronaut training (69.2) and medical imaging (66.4) which are close to the acceptable value of 70. In addition, the operational difficulty of the prototype in the confined cabin space of the airplane caused usability issues such as difficulty to properly recognize gestures in dark places. The student's SUS score for the player in all sessions are close to the acceptable score of 70 (see Table 2).

Instructional Design Methods (IDMs)

The trainer's ratings of IDMs across all three professional domains (see Section Use cases and application domains) is in general positive, ranking between 4 to 6, indicating positive acceptance of the implementation in all three professional domains (see Figure 8). Most IDMs such as Point of view *(M*=5.0, *SD*=1.414*),* Annotations *(M*=5.0, *SD*=1.414*),* Ghost Track *(M*=5.0, *SD*=1.414*)* etc. were rated above average by all the trainers in three sessions. Levene's test using the means (*p*=0.169) showed homogeneous variance between the three professional domains. In order to see if all the users of the three domains perceived the IDMs implementation equally which would hint that our prototype may be applicable across all 3 domains for training, we wanted to see if the differences between the results of the three domains were significant. Therefore, we conducted a MANOVA test and found a statistically significant difference in ratings between the three domains, F (24, 62) = 1.587, p = .075 ; *Wilk's* Λ = .384, *partial* $\eta 2$ = .381, mitigating the possibility that the results occurred by random occurrence. IDM Directed Focus was rated the lowest (*M*=4.25) across all the domains by the trainers. Details on the scores for each item can be found in Table 3.

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
		M (SD)	M (SD)	M (SD)	M (SD)
Directed Focus	DF1. I always knew where the next action happens	4.333 (0.730)	4.411 (1.175)	5.285 (1.112)	4.5 (0.707)
	DF2. I always knew where to stand and look	4.095 (1.374)	3.882 (1.053)	4.428 (0.786)	4.0 (0)
Highlight object of interest	HL1. I could always identify important objects	4.380 (0.864)	5.117 (0.992)	5.285 (0.487)	4.5 (0.707)
Point of View Video	POV1. Videos provided a trainer's point of view on the task	4.666 (0.966)	5.352 (0.861)	5.833 (0.983)	5.0 (1.414)
Cues and Clues	CUE1. The floating photos helped me understand what the task	4.714 (0.956)	5.470 (0.799)	5.714 (0.951)	5.0 (1.414)
Annotations	ANN1. The virtual sticky notes helped me identify important bits of information	4.523 (0.928)	5.470 (0.799)	4.714 (1.112)	5.0 (1.414)
Object Enrichment	OE1. The system provided related information on objects of importance	4.809 (0.980)	5.764 (0.752)	5.248 (1.272)	4.5 (0.707)

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
3D models and animations	ANI1. The 3D animations helped me to interpret complex concepts	4.476 (0.872)	5.058 (0.747)	5.142 (1.214)	5.0 (1.414)
Think aloud	TA2. I understood what to do when following the trainer's audio recordings	4.761 (0.943)	5.470 (0.717)	5.714 (0.951)	5.0 (1.414)
	TA1. Audio recordings provided an trainer's explanations	4.761 (0.889)	5.529 (0.624)	5.714 (1.380)	5.0 (1.414)
Contextual Information	CI1. The system provided information relevant to the current situation and process	4.666 (0.912)	5.352 (1.114)	6.0 (0.577)	4.5 (0.707)
Ghost track	GT1. I was able to identify the position and the spatial orientation of the recorded trainer	4.619 (0.920)	5.176 (0.727)	5.428 (0.975)	5.0 (1.414)

Table 3: Average trainer ratings of the IDM items

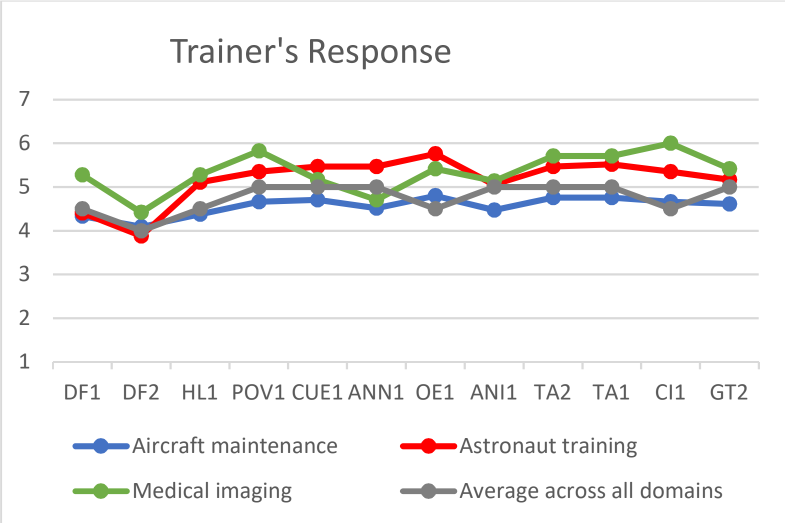

Figure 8. Trainer's response on the IDMs questionnaire

IMDs	Questions	Aircraft maintenance	Astronaut training	Medical imaging	Average across all domains
		M (SD)	M (SD)	M(SD)	M(SD)
Directed Focus	DF1. I always knew where the next action happens	5.58 (1.104)	3.454 (1.710)	4.410 (1.292)	6.0 (0)
	DF2. I always knew where to stand and look	5.35 (1.368)	3.454 (1.595)	4.205 (1.293)	5.5 (0.707)
Highlight object of interest	HL1. I could always identify important objects	4.088 (1.147)	5.272 (1.695)	5.410 (1.044)	6.0 (0)
Point of View Video	POV1. Videos provided a trainer's point of view	3.911 (1.815)	5.727 (1.279)	5.769 (0.916)	4.5 (2.121)
Cues and Clues	CUE1. The floating photos helped me understand what the task	2.617 (1.279)	5.772 (1.231)	5.666 (1.108)	3.5 (3.535)

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
Annotations	ANN1. The virtual sticky notes helped me find out what is important	4.970 (1.445)	5.545 (1.143)	5.564 (0.753)	6.5 (0.707)
Object Enrichment	OE1. The system provided general information for objects of importance	4.764 (1.327)	5.318 (1.170)	5.538 (0.853)	6.5 (0.707)
3D models and animations	ANI1. The 3D animations helped me to interpret complex models	5.058 (1.204)	5.272 (1.453)	5.589 (1.207)	6.0 (1.141)
Think aloud	TA2. I understood what to do when following the trainer's audio recordings	3.764 (1.102)	5.272 (1.241)	5.153 (1.159)	5.5 (0.707)
	TA1. Audio recordings provided a trainer's explanation	1.5 (1.022)	3.727 (1.723)	5.307 (1.217)	5.0 (1.414)
Contextual Information	CI1. The system provided information relevant to the current situation	3.941 (0.919)	5.818 (1.139)	5.512 (0.884)	5.0 (1.414)
Ghost track	GT1. I could identify the position and the spatial orientation of the recorded trainer	4.705 (1.030)	5.045 (1.252)	5.051 (1.050)	6.0 (0)

Table 4: Averaged student ratings of the IDM items

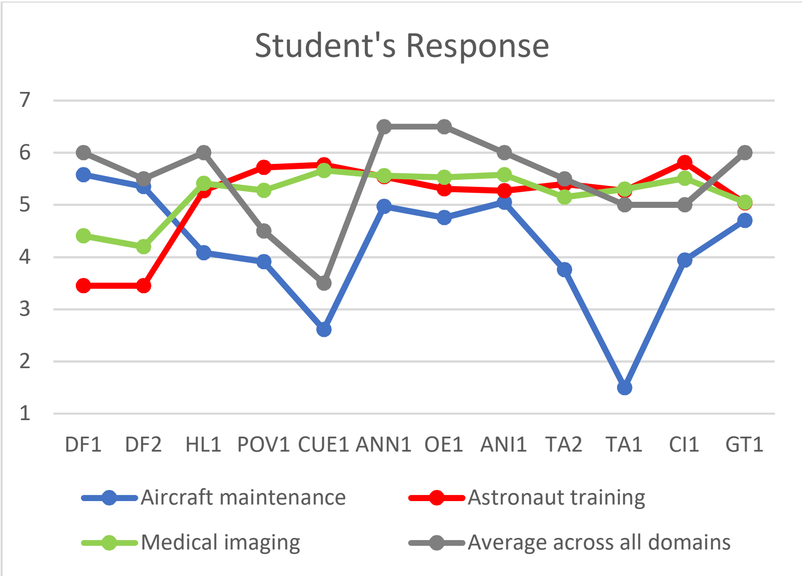

Figure 9: Student's response on the IDMs questionnaire

The overall students' ratings in all the three professional domains varied with scores ranging between 3 to 7 (see Figure 9). Levene's test using the means (p=0.07) showed a weak homogeneous variance between the three domains. This signifies that each domain perceives the IDMs differently. On conducting a MANOVA test, a statistically significant difference in ratings between the three domains, F (24, 162) = 14.097, $p < .0005$; *Wilk's* $\Lambda = .105$, *partial* $\eta 2 = .68$ was found mitigating the possibility that the results occurred by random occurrence. The IDM Directed focus (M=5.25) was rated the highest in the aircraft maintenance domain while Contextual information (M=5, SD=1.414) was rated the highest by the students in Astronaut training. IDM Cue and clue was rated the lowest by students in the aircraft maintenance with score of (M=2.617, SD=1.279) despite being rated above average by the students in astronaut training (M=5.772, SD=1.231) and medical imaging (M=5.666, SD=1.108). The IDM Think a loud (TA1: M=1.5, SD= 1.022) and (TA2: M=3.764, SD=0.102) was rated the lowest by the students in aircraft maintenance. Nonetheless, on average across all three domains, most IDMs were generally accepted with ratings above average. Details on the score can be found in Table 4. In conclusion the results show that the usability of the prototype is close to meeting the first hypothesis. Similarly, the general acceptance of most of the IDMs by the trainers and the students rating the IDMs above the average value of 4, show that our implementation of the IDMs in the prototype meets the definition of the authors. In addition, since it was rated by students across three professional domains, our third hypothesis on interoperability across the domains is also met except for a few IDMs in some domains.

5 Discussion

The trainer rating of IDMs has been above average in all three domains with only slight difference from the first study conducted in the aircraft maintenance domain to the second in astronaut training and the third medical imaging (see Figure 7). The aircraft maintenance trainers rated the IDMs lower than the other 2 domains which may have been due to the study being carried out inside the cabin of the plane with limited lights and moving space which required participants to crouch all the time and Hololens to often lose tracking of the environment. Increasing the recorders dependency on physical markers with Vuforia™ may help to create a more stable AR experience. Regardless, the core implementation of IDMs was generally accepted by all trainers across the three domains. Therefore, future iterations of the prototype will only focus on implementing more IDMs and improving the overall experience for the trainer. The recorder's usability was improved based on the observations in the first session which accounted for the positive usability ratings in later sessions. Prior to implementation of the graphical icons based navigation, the recorder implemented a text based navigation. The trainers were required to aim the cursor my moving his/her head onto the text and then make a tapping gesture in order to select the menu. This was inconvenient in such a confined space as many missed taps were performed by the trainers. Hololens display are by nature opaque to a certain degree and reading smaller texts were difficult. Therefore, Graphical icons were implemented to make navigation easier and more intuitive for the trainers. Menu's that required text were made bolder and larger to make tapping easier. The learning curve for the trainers who are unfamiliar with technology was higher we well. In order to support such trainers, built in tool tip or help is required which will shorten the learning curve and allow them to quickly adapt the technology in their traditional training classes. Navigation indicators to show that the recording is being performed by the prototype, was implemented with icons turning red during the recording process. The recorder will also implement voice based navigation to help improve the usability of the system.

In addition, instruction sheets for the participants were also improved based on the experience from the first session which contributed to the overall experience. We also observed that the trainers mostly used audio recordings to create learning content due to its simplicity. It should be noted that the trainers were mainly exposed to the recorder. The recorder only records data required for the IDMs. Thus, the IDMs themselves are not implemented except some such as directed focus and object enrichment which are useful to the trainer as well. Though the trainers were also exposed to the player briefly, it was done so to allow them to get an understanding of how the data they recorded was being used. In addition, questionnaire related to the IDMs were, by nature, more oriented to the player and the students. The individual ratings may have also been affected by the short time frame each trainer was given. The learning curve may have been higher due to the complexity of operating the recorder on top of the complexity posed by a new technology such as AR.

The students' average ratings for all IDMs also increased in later sessions despite the core implementation of IDMs in the player not being changed between the sessions. However, the usability of the recorder was improved which led to the trainers recording better content for the students which may have improved the overall perception of students. The highest perceived ratings of the students in each of the sessions varied according to the domain. The significance of each IDM may have varied according to its perceived usefulness for the domain of use. For example, the directed focus was rated significantly lower by students from astronaut training and medical imaging while being rated higher in aircraft maintenance session. Directed focus may have been perceived higher due to larger work area in aircraft maintenance where the IDM provided significant advantage. In other sessions, the students did not have to move from a single fixed position to perform the task. At the same time, IDMs such as object enrichment and 3D models and animations were rated in an equal manner among all three sessions which could potentially hint that such IDMs can be applicable across all these three domains. In Figure 2, the think aloud protocol was rated with a significant difference between the two sub-questions. TA2 asked if the participant understood reasoning behind the trainer's instructions as compared to TA1 which only asked if they understood what to do next. Trainers in aircraft maintenance were limited by constraints such as time and physical space which may have affected their explanations. Experts or trainers in this case, tend to underestimate how difficult a task can be for the students [Hinds 1999]. Trainers are also often unaware of all the knowledge behind their superior performance [Patterson 2010] and thus may omit the information an student would find valuable [Hinds 2001]. The largest pool of trainers in the aircraft maintenance session had limited time which did not allow each step to be comprehensively elaborated. Furthermore, it may have been due to the instructions not being explicit to the trainer, which was improved over the upcoming sessions. This is reflected in Figure 8, where average ratings for the think-aloud protocol has improved in the latter sessions. IDM Cues and clues was rated the lowest across all three sessions, due to significant low ratings in the aircraft maintenance. It is unclear now as to why it was rated so and needs further analysis.

To summarize, this paper reports the first user study of the prototype designed to support expertise development utilizing the recorded trainer performance data. The prototype is developed as a part of design based research project with forthcoming iterations in the future. Performing this study has provided us with a baseline for the measure of usability and a measure of proper implementation of IDMs. Combining various IDMs to enable support for different professional domains can generate many risk and challenges. Implementation of many IDMs may lead to increased complexity in the software and risk that each IDM implementation may fail to fulfil their purpose due to overhead in mixing various IDMs together. It is crucial to explore different approaches to design the system that reduces the learning curve, increases usability and overall achieves all the benefits of each implemented IDM.

6 Limitations and Future work

The implemented IDMs need to be better represented by the system before we can measure the learning outcome provided by the system. Based on observations, the recorder must implement other functionalities in a more intuitive manner reducing the learning curve for the trainers. This could otherwise limit the results of the future studies and the IDMs available to the students as the player depends on the recorded trainers' data. Both the students and trainers might also have been overwhelmed learning the new technology and range of functionalities implemented in the prototype in such a short time. To account for this, the system needs to be more intuitive. In addition, proper instructions can also be provided to the users along with more time allocated to each user to use the prototype. Due to the short time provided to the both the students and the trainers during the sessions, we were not able to collect much required qualitative data. Future studies may be focused more on smaller groups with more time for exposure. In addition, there are many difficulties trainers face to adapt the system in their regular training sessions as observed during the sessions. The system needs to support this transition to the best possible manner. It must also be complemented by proper instructions and training to support this transition.

Finally, more IDMs need to be implemented to support the domains more concretely. IDMs whose implementation were rated poorly will be further analysed and discussed with the trainers and the students to improve their implementation. The prototype used in this study was a linear system with minimal feedback being provided to the user. Proper feedback mechanisms will be implemented to enhance the usability and intuitiveness of the system. The usability of the system itself is not yet in an acceptable range. AR based usability guidelines will be further closely integrated to improve the usability in the system. Audio based interaction and proper user interface design to ease the learning of the system are some of the aspects that need improvement.

References

[Bacca 2014] Bacca, J., Fabregat, R., Baldiris, S., Graf, S., Kinshuk, Fabregat, R., Graf, S.: 'Augmented reality trends in education: A systematic review of research and applications'; Educational Technology & Society, Vol. 17, No. 4 (2014), pp. 133–149. Retrieved from http://www.jstor.org/stable/jeductechsoci.17.4.133

[Bangor 2009] Bangor, A., Kortum, P., Miller, J.: 'Determining what individual SUS scores mean: Adding an adjective rating scale'; Journal of Usability Studies, Vol. 4, No. 3 (2009), pp. 114–123. https://doi.org/66.39.39.113

[Carey 2014] Benedict Carey: 'How Do You Get to Carnegie Hall? Talent - NYTimes.com'; (2014). Retrieved from https://mobile.nytimes.com/2014/07/15/science/which-matters-more-talent-or-practice.html?referrer=

[Brooke 2013] Brooke, J.: 'SUS: a retrospective'; Journal of Usability Studies, Vol. 8, No. 2 (2013), pp. 29–40. Retrieved from http://dl.acm.org/citation.cfm?id=2817913

[diSessa 2004] diSessa, A. A., Cobb, P.: 'Ontological Innovation and the Role of Theory in Design Experiments'; Journal of the Learning Sciences, Vol. 13, No. 1 (2004), pp. 77–103. https://doi.org/10.1207/s15327809jls1301_4

[Drljevic 2017] Drljevic, N., Wong, L. H., Boticki, I.: 'Where does my Augmented Reality Learning Experience (ARLE) belong? A student and teacher perspective to positioning ARLEs'; IEEE Transactions on Learning Technologies (2017), pp. 1–1. https://doi.org/10.1109/TLT.2017.2690426

[Ericsson 2006] Ericsson, K. A., Charness, N., Feltovich, P. J., Hoffman, R. R.: 'The Cambridge handbook of expertise and expert performance'; Cambridge Handbooks in Psychology. Cambridge University Press (2006).

[Ericsson 2007] Ericsson, K. A., Prietula, M. J., Cokely, E. T.: 'The Making of an Expert. (cover story)'; Harvard Business Review, Vol. 85, No. 7/8 (2007), pp. 114–121. https://doi.org/Article

[Gladwell 2008] Gladwell, M.: 'Outliers: The Story of Success.'; Academy of Management

Perspectives. New York: Brown and Company (2008).

[Hinds 1999] Hinds, P.: 'The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance.'; Journal of Experimental Psychology: Applied, Vol. 5, No. 2 (1999), pp. 205–221. https://doi.org/10.1037/1076-898X.5.2.205

[Hinds 2001] Hinds, P. J., Patterson, M., Pfeffer, J.: 'Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance.'; Journal of Applied Psychology, Vol. 86, No. 6 (2001), pp. 1232–1243. https://doi.org/10.1037/0021-9010.86.6.1232

[Jarodzka 2017] Jarodzka, H., Holmqvist, K., Gruber, H.: 'Eye tracking in educational science : Theoretical frameworks and research agendas'; Journal of Eye Movement Research, Vol. 10, No. 1 (2017), pp. 1–18. https://doi.org/10.16910/jemr.10.1.3

[Olwal 2008] Olwal, A., Gustafsson, J., Lindfors, C.: 'Spatial augmented reality on industrial CNC-machines'; In Proceedings of SPIE (Vol. 6804) (2008). https://doi.org/10.1117/12.760960

[Patterson 2010] Patterson, R. E., Pierce, B. J., Bell, H. H., Klein, G.: 'Implicit Learning, Tacit Knowledge, Expertise Development, and Naturalistic Decision Making'; Journal of Cognitive Engineering and Decision Making, Vol. 4, No. 4 (2010), pp. 289–303. https://doi.org/10.1177/155534341000400403

[Schneider 2017] Schneider, J., Börner, D., van Rosmalen, P., Specht, M.: 'Presentation Trainer: what experts and computers can tell about your nonverbal communication'; Journal of Computer Assisted Learning, Vol. 33, No. 2 (2017), pp. 164–177. https://doi.org/10.1111/jcal.12175

[Wu 2013] Wu, H. K., Lee, S. W. Y., Chang, H. Y., Liang, J. C.: 'Current status, opportunities and challenges of augmented reality in education'; Computers and Education, Vol. 62 (2013), pp. 41–49. https://doi.org/10.1016/j.compedu.2012.10.024