Workshop Proceedings of the 11th International Conference on Intelligent Environments
D. Preuveneers (Ed.)

doi:10.3233/978-1-61499-530-2-236

The GhostHands UX: telementoring with hands-on augmented reality instruction

Giuseppe SCAVO [a], Fridolin WILD [a] and Peter SCOTT [a]
[a] *Knowledge Media Institute, The Open University, United Kingdom*

Abstract. Augmented Reality is providing a whole new set of possibilities to improve proficiency and safety on the workplace and reduce costs during training, in particular in manufacturing and automotive industries. Among the applications developed to support workers on the shop floor, telementoring plays a leading role as a demonstrator of the capabilities of AR in industry. This paper explores the user experience of telementoring via GhostHands - the use of a virtual model of a mentor's real hands, appearing in a workers full AR view of a work task. The mentor can see the workers view 'through their eyes' and can place their GhostHands into that view. Both worker and mentor appear to cope with this 'ghost assisted' experience very well, and recognise a sense of joint accomplishment from the cues that the hands embody in completing a task.

Keywords. Augmented Reality, Telementoring, User Experience, Instruction

1. Introduction

Over the past ten years, one key driver behind the development of Augmented Reality (AR) technologies has been industrial and manufacturing contexts. In these environments, there are many advantages to overlaying real-time contextual information directly into the worker's field of view of the workplace, such as a increasing learning and work efficiency and improved awareness and safety. Considering the recent expansion of the market for "full vision" AR glasses, AR applications in the industrial context can now in principle be completely hands-free. One full vision application in particular, that is capturing the attention of both research community and industrial partners is 'telementoring'. The possibility to connect with a remote expert mentor who can support or train workers during their actual working performance is a very powerful idea.

Although these new channels of communication represent a great advancement in industry, they are usually a poor substitute for the physical real-time presence of a mentor. One of the most important challenges is that a physical human body in a physical workplace "embodies" many subtle learning cues. A worker can learn very effectively via the non-verbal observation of a mentor performing a manual task via subtle cues 'embodied' in eg. the position of the mentors hands, cued by the hands' position, angle, thrust, or safe movement. In this paper, we present GhostHands, a novel interaction concept that tries to fill this gap by overlaying on the worker's field of view the virtual hands of an expert – modelled in real-time, from the position of that expert's remote, real hands. The purpose of this paper is not to evaluate the actual effectiveness of the prototype application in terms of quantitative measurements of task completion time or error-rate trends, but to provide insight into the experience of users, on both ends of this interaction link. The user

experience strongly affects the effective adoption of a new technology and how it will impact, in this case, on work performance. The results of this study will be used to inform the design in the next development iterations of this concept.
Section 2 identifies background and rationale behind GhostHands. In Section 3 the application and its implementation are described in detail as well as the evaluation protocol. The results of the evaluation are reported in Section 4 and discussed in Section 5.

2. Related research

Through AR-enabled devices the worker can contact an instructor during difficult situations, such as wicked problems or complex tasks that require specialist knowledge, critical procedures involving safety concerns, or variations of configurations of machines that the worker has no direct knowledge of, hence requiring cross-training. In particular in non-standard cases, packaged AR applications fail to provide the needed support for effective problem resolution and remote collaboration is essential for the solution of the problem. Remote collaboration also remarkably improves proficiency of workers during manual labor [6]. In the past, many works have tried to explore the possibilities of AR-enabled remote mentoring. Zhong et al. [2] propose a prototype of an AR collaborative training application that creates a "shared" workspace in which the trainer can manipulate virtual objects overlaid on the trainee's field of view. Interaction between users, however, was in this case limited to simple real-time audio conversation. Bottecchia et al. [3] propose an AR system for tele-assistance based on a new interaction protocol, that allows the instructor to point at objects, outline particular parts of it and apply 3D overlaid animations. Although granting to the expert a good level of interaction with the worker's environment, this application is based on the assumption of having 3D models of all the machines and tools the worker is going to operate with. This assumption is not applicable in real-life use case, where third-party supplying companies are often not allowed to release such information. Similarly, a body-worn camera/laser pointer - named WACL - is proposed in [4]. WACL allows the expert to point and name objects to accelerate the collaborative work. An alternative approach is taken in [5]. In this platform, the "Adaptive Visual Aids" are visual post-its in the form of shapes, text or drawings that can be superimposed on the worker's field of view independently from the object position and are shareable among workers in different tele-consultancy sessions. All the systems described suffer the following limitation: the identification of the worker with the expert. In a side-by-side working situation, the worker can see what the expert's movement and gestures are, and acquire visual knowledge of the task by simply watching and then repeating. The current state of the art of AR applications does not offer an appropriate replacement for this learning channel. Moreover, it is demonstrated that workers perform better when the expert is physically present because of "their ability to gesture in the space to ground the conversation more efficiently" [1]. Gestures play a fundamental role in timing the conversation and reducing the stress caused by communication synchronisation. In addition, to facilitate the interaction, spatial indications such as "here" and "there", or "this" and "that" are indispensable for an effective and comfortable discussion [7].

3. GhostHands

We propose a new interaction modality for AR-enabled telementoring: GhostHands. The prototype GhostHands AR application allows telementors to remotely stream a realistic

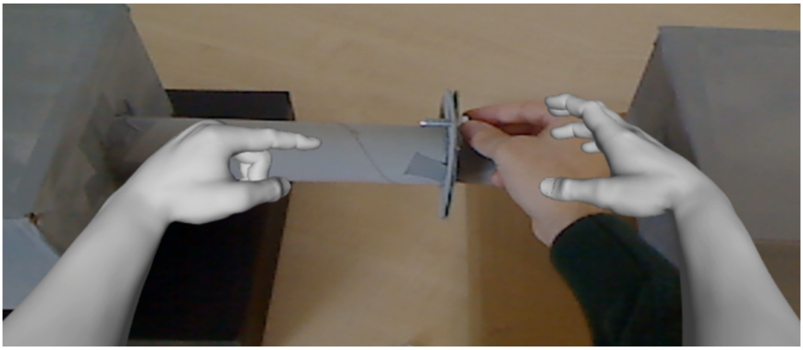

Figure 1. A screenshot of the GhostHands application.

model of their forearms+hands directly onto a worker's field of view. In addition to the shared visual space and the conversational audio communication via real-time video and audio streams, the mentor is able to interact with the worker through the direct use of hand movements and gestures.

GhostHands is composed of two separate intercommunicating applications. The expert-side application shows on a display the live video stream of the worker's environment. By means of an infrared sensor, the application captures the mentor's hand movements and reproduces them in real-time on the display using 3D models of human-like hands. The 3D models are overlaid on the video feed, so that the expert can constantly observe how their "ghost hands" are spatially positioned with respect to the incoming video stream. The mentor has complete freedom of movement and is able to perform a very wide variety of gesture within the worker's view, moving both their hands with six degrees of freedom. On the worker-side application, a head-mounted camera captures the live video stream that will be sent to the mentor side's counterpart, while a display shows to the worker the 3D modelled remotely-controlled hands in situ.

Using GhostHands the expert can virtually interact with the worker in a reaching "over-the-shoulder" way, performing many kind of actions:

- pointing at objects in the workers environment and naming them without the need of graphic pointers;
- signalling a danger;
- showing how to properly interact with objects performing the correct movements, demonstrating the proper hand's pose, angle and speed;
- performing gestures while talking.

The main advantage of GhostHands lays in the possibility for the worker to instantly mimic the expert's movement. The worker can virtually "fit" his/her hands in the expert's ones, immediately reproducing pose and movement that overlaid hands are performing. This produces a great benefit in terms of both kinaesthetic learning and procedure ergonomics. The virtual guidance provided by the expert through GhostHands can partially compensates his/her physical absence. In addition, with GhostHands the worker can simultaneously watch the hands instructing and perform the task, fitting his/her hands into the mentors ones like virtual gloves.

This interaction modality considerably benefits also situations where the workplace safety is at risk. Imagine a scenario in which a worker needs to face an unexpected situation of

danger. A machine processing toxic chemical compounds suddenly stops working with the risk of material leakage and none of the workers in the facility knows the procedure. Using GhostHands a remote expert can supervise the procedure instructing a worker on actions to perform, maintaining the safety level high and avoiding the risk for the worker to incur in physically harming accidents.
The communication signalled by hand gestures is also highly relevant. Being able to perform spatial gestures greatly reduces the need for extensive and verbose descriptions of the environment around the worker, thus speeding up the communication between the two sides. Gesticulation is also a natural synchronisation interface in a face-to-face human conversations. Hence, the absence of this visual channel can cause stress and frustration. GhostHands offers to fill these gaps providing a natural, virtual interface to compensate the physical absence of the expert in the workers environment.

3.1. Implementation

Both expert-side and worker-side applications have been built using the Unity3D game engine. To capture the experts hand movement, the capabilities of the Leap Motion sensor has been exploited. This sensor projects infrared light from a USB-tethered device, capturing hand movements. The expert user hovers their hands over the device in order to use it. The Leap Motion SDK provides a WebSocket server that streams captured data to our the GhostHands service. To reproduce hand movements on the other end of the connection, this stream has been redirected via a Node.js server that bridges and regulates the communication between worker-side and expert-side application.

3.2. User eXperience evaluation

In the majority of the studies involving AR-enabled technologies and applications, research has so far primarily focused on investigating the advantages of AR in terms of performance improvement or, in some cases, on identifying issues related to usability or cognitive stress [8,9]. Generally, the AR research community lacks understanding of the aspects of user experience for AR demonstrators [10]. UX insights are, however, crucial if the community aims at delivering well-accepted services. For this reason, we considered necessary to perform a user-centered evaluation of GhostHands. The results of this evaluation will be fed in the following development iterations and will serve as a benchmark relative to user requirements and expectations. The overall objective of this work is to obtain insights on how the introduction of a natural communication interface, such as hands gestures, is perceived by users when channeled into a new technology like AR. To achieve this, some relevant questions need to be answered:

- How do GhostHands improve the worker's learning engagement?
- How do GhostHands affect the level of confidence of both the instructor and the worker?
- To what extent do users perceive the task facilitated by the usage of GhostHands?

User experience concepts are often neglected in interaction design. Disregard for human factors when designing a new interaction model, as GhostHands is, might result in side effects due to the failure of engagement into the new technology (e.g. visceral repulsion towards the technology and decrease in productivity). On the other hand, positive experiences and emotions related to the product improve users focus, facilitating and increasing well-being [11].

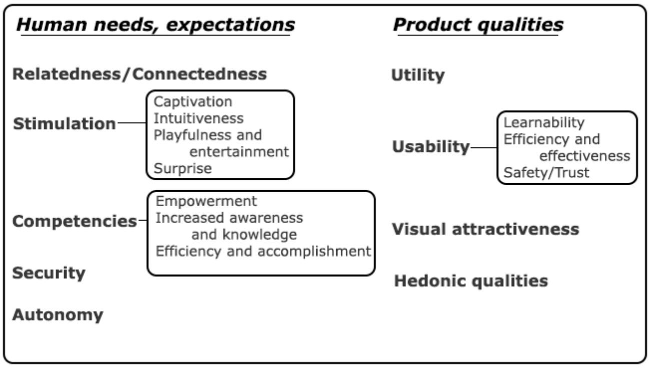

Figure 2. The UX evaluation framework used in this study

3.3. UX framework

The importance of User eXperience has been investigated and thoroughly established in the research community during the last fifteen years [12,13]. However, given the subjective nature of the topic, no standard framework has arisen as generally accepted by the community. Therefore, many of the frameworks proposed are specific to some research areas and tend to evaluate UX of a specific category of products. Although these frameworks differ for methodologies used, common traits can be identified. The most important of all is the categorisation of factors influencing the user experience of the product in two different categories: product qualities and human needs/expectations. The evaluation framework adopted in this paper strongly relies upon this subdivision, taking as model the framework proposed by Schulze and Kromer [14]. In their work Schulze and Kromer propose a framework for UX evaluation of online interactive products. The UX is the result of two influencing factors: basic human needs, such as relatedness, popularity or competency, and product qualities, such as usability or visual attractiveness.
Product qualities have been extensively studied. In [15], the most significant product quality measures are selected and commented. For what concerning human needs and expectation instead, the subjectivity of the matter imposes deeper investigation in the application context, in this case AR. Scarce consideration has been given to AR-related UX, and this is demonstrated by the absence of an evaluation framework specific to AR. Olsson et al. [10] identify meaningful components for the expected user experience of AR services. The influencing elements selected are mostly an operationalisation into an AR context of the basic human needs described by Schulze. The aforementioned frameworks were exploited to select a number of measures that this evaluation will consider. The resulting framework is shown in Figure 2.

3.4. The experiment

To evaluate the experience of GhostHands, a lab experiment was prepared. The setting consisted of 6 pairs of participants cooperating in pairs to perform a sequence of tasks. For each pair of participants, one of them impersonated the instructor, while the other was participating as trainee. The instructor was briefed before the experiment on the tasks to perform and the terminology to use to describe object. In order to simulate a factory-like scenario, the tasks were designed around piece of machinery currently used in industry that were replicated using cardboard. Three tasks of increasing complexity were designed to be performed in sequence with short rest intervals:

1. Rotor disc servicing: a rotor disc mounted on a shaft had to be rotated to compensate for wear and tear. (19 steps)
2. Motor/pump alignment: the pipe connecting motor and pump was misaligned and required realigning in order to prevent malfunction. (24 steps)
3. Thread feeder unit replacement: a very delicate piece of machinery that holds thin threads needed to be replaced through a complex and strict series of steps. (33 steps)

Each session, the two participants were separated by a barrier that prevented them to see each other. The trainee was equipped with headphones and a head-mounted camera, used to capture the video stream sent back to the expert. In front of him/her, the piece of machine to operate on was laying on a table. In background, a large screen showed to the trainee the video stream of what was being captured by the camera with the instructor's GhostHands superimposed. The instructor instead, was equipped with a laptop and the hands tracking sensor.
An open answer post-task questionnaire was presented to the participants after the experience. The questionnaire included questions about their experience and emotions regarding specific aspects of the experiment.

4. Results

As a general trend, the experience resulting from the experiment was of empowerment and safety. Trainees generally felt more self-confident knowing that an instructor was guiding them and teaching how to properly perform the task. This can also be partially caused by simply having the instructor guiding "over-the-shoulder", but some participants explicitly stated that GhostHands was particularly useful when the task was not obvious. "*It got me out of a few instances where I was stuck*". "*GhostHands allowed me to feel more 'safe' about the actions I was assigned to. As a learner I felt more empowered, since my confidence level was higher*". However, several participants that participated as instructors raised concerns about the precision of the application. During the gestures performance, the sensor can lose track of hands, especially if the user is not familiar with such a technology. That results in the hands model not accurately reproducing the gesture and the instructor needing to recalibrate the hands on the sensor. In the following sections findings will be presented with more detail.

4.1. Product qualities perception

All the participants found that this interaction modality generally sped up the instructions. "*Even though the tasks were relatively simple, I imagine it would have taken me more time to understand and carry out the tasks without the GhostHands*". Instructors perceived the interaction as natural and intuitive, using hands as if physically present in the trainee's space. For many of them being able to perform spatial gestures was greatly convenient. "*When struggling with or forgetting some name of the component, you just point at it. The same applies when you cannot find the proper verb or description*". However, the frequency with which instructors took advantage of the possibility to guide the trainee actively showing the proper movement to perform, was lower than expected. This is mostly due to the lack of confidence in using the sensor and the imprecision in hands tracking.

All the instructors considered that the situation in which hands were mistracked, affected negatively the experience. "*I would trust it if the imprecision is somehow addressed*". Also trainees in particular situations that required higher precision, considered wearing the camera somewhat uncomfortable. "*Unfortunately the precision of the instructions is heavily affected by the movements of the worker's head. He must be aware of that and avoid shaking his head*". In cases when instructors reached a confidence level high enough to perform gestures naturally, trainees felt safely guided through the task. "*GhostHands was as if I had a pair of expert hands taking mine and guiding me through the entire process*". Overall, both sides of the interaction experienced a significant improvement in perceived efficacy. Feedback on efficiency and effectiveness were extremely positive. "*When it comes to physical movements, written instructions and static pictures aren't very helpful and often misleading. Have you ever tried to follow origami instructions?*". Some participants, instructors in particular, suggested the integration with more visual information channels, such as arrows and shapes to improve pointing at small objects. For what concerning visual attractiveness, white plastic-like hands were considered suitable and "*appropriate for this industry-oriented tasks*". The majority of the participants were skeptical when asked if there was a possibility to improve the 3D hand model. In their opinion, models that look too realistic would feel uncanny, while it would be difficult to adapt to models that do not accurately represent human hands.

4.2. Human needs, expectations, visceral feelings

Feelings of empowerment and efficiency arise throughout participants' answers, as highlighted in Section 4.1. During the execution, instructors declared that operating with the tracking sensor became more and more playful and engaging as they mastered it. Occasionally they felt frustrated because of the poor tracking. "*I had great fun playing with the GhostHands and by the end of the experiment, I had started to feel confident. There were slight moments of frustration due to losing the GhostHands tracking/motion*". Except for the tracking issues, none of the users expressed negative feelings about the intuitiveness and naturalness of the interaction.

In some cases the playfulness of the interaction was also associated with the connection with the other side of the communication. The instructor congratulated the worker with a "thumbs-up" gesture, or they interacted with an "high five". "*I think it can help in improving team-work because it produces a sense of 'shared accomplishment'*". A certain number of concerns were expressed with reference to feelings of autonomy. The strong guidance provided by the instructor could interfere with the active learning process of the trainee. This can result in distraction and passive attitude towards the instructor. "*I felt that sometimes the worker may rely completely on the instructor being like a robot executing the instruction without really learning*".

5. Discussion

The aim of this work is to present GhostHands and evaluate this new interaction modality from a UX perspective. Users generally perceived the experience as greatly stimulating and with a strong sense of connectedness and playfulness, hence improving engagement. Instructors felt closer to their trainees while trainees felt safe and guided. This resulted in increased self-confidence and sense of empowerment, two feelings extremely important during the learning process. Results show how important is precision of the application for instructors. If the application fails, even only occasionally, at representing gestures and movements in real-time, hands models are perceived as uncanny. This produces a feeling of frustration that can discourage users from taking advantage of all the capabilities of such interaction. However, the major advantage of this cooperation channel lays in the naturalness and intuitiveness of the interface. Technology seemed to be completely transparent for users. All the participants learnt how to interact with the application in matter of seconds, and some mastered it in minutes, despite the fact that they were completely unfamiliar with the technology before the experiment. Therefore, one can assume that their capability to deal with the inaccuracies of the sensor will improve with learning.

All in all, GhostHands introduces a new channel of communication and learning perceived as natural and connecting, that can be integrated with the current AR-enabled interaction modalities to improve effectiveness and precision.

Our iteration in this research will focus on improving the stability and precision of tracking which is crucial for the final users. Concerns about effectiveness in learning will be addressed through a series of quantitative studies, and the development agenda also includes further application deployment for smartphones, tablets and smart glasses.

References

[1] Kraut, Robert E., Susan R. Fussell, and Jane Siegel. "*Visual information as a conversational resource in collaborative physical tasks.*" Human-computer interaction 18.1 (2003): 13-49.

[2] Zhong, Xiao Wei, Pierre Boulanger, and Nicolas D. Georganas. "*Collaborative augmented reality: A prototype for industrial training.*" 21th Biennial Symposium on Communication, Canada. 2002.

[3] Bottecchia, Sèbastien, Jean-Marc Cieutat, and Jean-Pierre Jessel. "*TAC: augmented reality system for collaborative tele-assistance in the field of maintenance through internet.*" Proceedings of the 1st Augmented Human International Conference. ACM, 2010.

[4] Sakata, Nobuchika, Takeshi Kurata, and Hideaki Kuzuoka. "*Visual assist with a laser pointer and wearable display for remote collaboration.*" (2006).

[5] Webel, Sabine, et al. "*An augmented reality training platform for assembly and maintenance skills.*" Robotics and Autonomous Systems 61.4 (2013): 398-403.

[6] Kraut, Robert E., Mark D. Miller, and Jane Siegel. "*Collaboration in performance of physical tasks: Effects on outcomes and communication.*" Proceedings of the 1996 ACM conference on Computer supported cooperative work. ACM, 1996.

[7] Bolt, Richard A." *'Put-that-there': Voice and gesture at the graphics interface*". Vol. 14. No. 3. ACM, 1980.

[8] Drascic, David, and Paul Milgram. "*Perceptual issues in augmented reality.*" Electronic Imaging: Science & Technology. International Society for Optics and Photonics, 1996.

[9] Dünser, Andreas, Raphal Grasset, and Mark Billinghurst. "*A survey of evaluation techniques used in augmented reality studies.*" Human Interface Technology Laboratory New Zealand, 2008.

[10] Olsson, Thomas, et al. "*Expected user experience of mobile augmented reality services: a user study in the context of shopping centres.*" Personal and ubiquitous computing 17.2 (2013): 287-304.

[11] Desmet, Pieter, and Paul Hekkert. "*Framework of product experience.*" International journal of design, 1 (1) 2007 (2007).

[12] Hassenzahl, Marc, Sarah Diefenbach, and Anja Göritz. "*Needs, affect, and interactive products - Facets of user experience.*" Interacting with computers 22.5 (2010): 353-362.

[13] Hassenzahl, Marc. "*The thing and I: understanding the relationship between user and product.*" Funology. Springer Netherlands, 2005. 31-42.

[14] Schulze, Katrin, and Heidi Krömker. "*A framework to measure user experience of interactive online products.*" Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research. ACM, 2010.

[15] Bevan, Nigel. "*Classifying and selecting UX and usability measures.*" International Workshop on Meaningful Measures: Valid Useful User Experience Measurement. 2008.