Assembly Training Using Commodity Physiological Sensors

Melynda Hoover, Anastasia MacAllister,
Joseph Holub, Stephen Gilbert, Eliot Winer
Iowa State University
Ames, Iowa
mthoover@iastate.edu, anastac@iastate.edu,
jholub@iastate.edu, gilbert@iastate.edu,
ewiner@iastate.edu

Paul Davies
The Boeing Company
St. Louis, Missouri
Paul.r.davies@boeing.com

ABSTRACT

Wearable technology is a thriving industry with projections for continued growth in the next decade and numerous unexplored applications. The U.S. Military has been on the forefront of this technology by supporting the research and development of these devices for today's warfighters. Smartwatches with sensors that detect physiological responses, like heart rate, have particularly interesting applications to warfighters. These devices have the potential to detect user stress during many different tasks from field operations to maintenance. Specifically, this paper will analyze the use of commodity sensors for evaluating and improving Augmented Reality (AR) work instructions. These AR work instructions have been shown to improve accuracy and efficiency in assembly tasks, which is crucial to the maintenance of military fleets. The study described in this paper compares two different wrist sensors, the Apple Watch and the Empatica E4. The Apple Watch is a popular, low-cost commodity wrist sensor, while the Empatica E4 is a higher cost, medical grade sensor. Participants wore both sensors while assembling a mock aircraft wing using work instructions delivered through an AR system. During the study, data such as errors, completion time, and several self-reported measures were recorded in addition to heart rate. After the study was completed, the heart rate data was extracted from the devices and analyzed. The results showed that the Apple Watch was less reliable because of its lower sample rate and gaps in data possibly due to user hand movement. Alternatively, the Empatica E4 was able to identify heart rate differences in steps of high and low difficulty with a lower standard deviation within steps. Based on these results, it was determined that the Empatica E4 was a more viable sensor for evaluating AR work instructions and that commodity sensors most likely need improvement before use in an industrial / military setting.

ABOUT THE AUTHORS

Melynda Hoover is a graduate student in Mechanical Engineering at Iowa State University's Virtual Reality Applications Center. She is working on evaluating the potential applications of wearable heart rate sensors to aid in developing Augmented Reality work instructions.

Anastacia MacAllister is a graduate student in Mechanical Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center. She is working on developing Augmented Reality work instructions for complex assembly and intelligent team tutoring systems.

Joseph Holub is a graduate student in Computer Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center.

Stephen Gilbert, Ph.D., is an associate director of the Virtual Reality Applications Center and assistant professor of Industrial and Manufacturing Systems Engineering at Iowa State University. His research interests focus on technology to advance cognition, including interface design, intelligent tutoring systems, and cognitive engineering. He is a member of IEEE and ACM and works closely with industry and federal agencies on research contracts. He is currently PI on a project supporting the U.S. Army Research Laboratory STTC in future training technologies for teams.

Eliot Winer, Ph.D., is an associate director of the Virtual Reality Applications Center and associate professor of Mechanical Engineering and Electrical and Computer Engineering at Iowa State University. He is currently co-leading an effort to develop a next-generation mixed-reality virtual and constructive training environment for the U.S. Army. Dr. Winer has over 15 years of experience working in virtual reality and 3D computer graphics technologies on sponsored projects for the Department of Defense, Air Force Office of Scientific Research, Department of the Army, National Science Foundation, Department of Agriculture, Boeing, and John Deere.

Paul Davies is an electrical engineer specializing in digital signal processing, and works in the Production Systems Technology group in Boeing Research & Technology. Since joining Boeing in 2003 he has supported the Advanced Tactical Laser, Homeland Security & Services, Delta II and B1B programs in addition to multiple IRAD and CRAD projects in Signal Processing, Augmented Reality and Machine Vision. He currently develops technology for Augmented Reality in manufacturing and investigates new methods of person-machine interaction for technician support. Paul received a BS degree in Electrical Engineering from Rochester Institute of Technology in May 2004, and a MS degree in Electrical Engineering from California State University Long Beach in May 2008.

Assembly Training Using Commodity Physiological Sensors

Melynda Hoover, Anastasia MacAllister,
Joseph Holub, Stephen Gilbert, Eliot Winer
Iowa State University
Ames, Iowa
mthoover@iastate.edu, anastac@iastate.edu,
jholub@iastate.edu, gilbert@iastate.edu,
ewiner@iastate.edu

Paul Davies
The Boeing Company
St. Louis, Missouri
Paul.r.davies@boeing.com

INTRODUCTION

In a forecast by the International Data Corporation (IDC), an organization which specializes in consumer technology market predictions, the shipment of wearable sensors is predicted to reach 214.6 million units by the year 2019, with the majority of these new sensors being smartwatches (IDC, 2015). Specifically, current wrist worn wearable sensors like the Apple Watch and Fitbit represent a large portion of a booming wearable sensor market. The military has been quick to recognize the potential applications of wearable sensor technology. In 2015, they awarded a $171 million partnership with industry and academia to develop wearable sensors that can be used by today's warfighters (U.S. Department of Defense, 2015).

With the growing accessibility and popularity of smartwatches, they are a logical choice for monitoring stress or cognitive load during the daily activities of warfighters. The convenient location of smartwatches on a peripheral pulse point of the body allows photoplethysmography (PPG) sensors to pull data from a human seamlessly. PPG is a method of measuring blood volume pulse (BVP) using light reflectance and photodetectors on the surface of the skin at one of several different pulse points. This method is advantageous and often used in smartwatches because it is relatively inexpensive and non-invasive compared to a traditional electrocardiogram (ECG). PPG sensors can be applied to many different pulse points on the body, including the wrist, finger, and ear (Allen, 2007). Several successful studies point to PPG sensors, such as those used in smartwatches, as an accurate indicator for stress (Lyu et al., 2015; Yoo & Lee, 2011). This stress sensing capability is one with many military applications that have yet to be explored. However, some challenges, such as battery life and computing power, may still require improvement if smartwatches are to be used for continuous physiological monitoring (Rawassizadeh, Price, & Petre, 2015).

One prospective application of wearable physiological sensors, which the U.S. Air Force is currently developing, is a system called BATDOK. This system allows medics in the field to continuously monitor the vital signs of injured warfighters and treat them more efficiently (Hymel, 2015). Another application, and the one which will be focused on throughout this paper, is the evaluation of Augmented Reality (AR) systems. An AR system is defined in Azuma's seminal paper as one which "allows the user to see the real world, with virtual objects super-imposed upon or composited with the real world. Therefore, AR supplements reality, rather than completely replacing it" (Azuma, 1997). Much research has been conducted which proves that employing AR systems to assembly work instructions has advantages such as increased efficiency and accuracy (Gavish, Gutiérrez, & Webel, 2015; Henderson & Feiner, 2009; Richardson et al., 2014). These qualities of AR are especially advantageous to the military because of the numerous warfighting vehicles and other equipment that they maintain. However, fine tuning these instructions to the needs of warfighters and maintenance personnel requires more detailed research. In a comparison of 165 different publications that evaluated AR systems via user studies, the majority used evaluation techniques such as completion time, error counts, scores, number of actions, and user-reported measures (Dünser, Grasset, & Billinghurst, 2008). While observational methods are capable of evaluating an AR interface as a whole, they are not very efficient at pinpointing the steps that cause stress to a user. This perceived stress is a result of anxiety or fear of failure which often causes physiological arousal (Boff, Kaufman, & Thomas, 1986). These physiological responses could be triggered by AR technology in several different ways. For example, confusing instructions or an overwhelming amount of information could cause stress to the user, especially if the user is unfamiliar with the AR platform. Self-reporting methods can collect this data, but not without introducing user bias or the possibility of misreporting. By continuously monitoring user heart rate during a study, using a PPG wrist sensor, it may be easier to discover heart rate trends and pin point the tasks for which AR is less effective. If this method is successful, it could help to evaluate the AR work instructions while eliminating the user bias of self-reporting techniques.

For the study described in this paper, two different PPG enabled watches were chosen. The first was a medical grade sensor called the Empatica E4. Medical grade sensors are considered more reliable and accurate and must often meet one or more regulatory compliances set forth by the medical community in order to ensure safety to the consumer (World Healthcare Organization, 2003). The Empatica E4 was chosen because it has been studied and verified in several different medical applications such as seizure reporting (Bidwell, Khuwatsamrit, Askew, Ehrenberg, & Helmers, 2015), stress and sleep monitoring (Muaremi, Arnrich, & Tröster, 2013). The rigorous testing and verification of the Empatica E4 sensor indicate it to be a reliable measure of heart rate. In addition, this sensor is costly (i.e. ~$1500-$2000 per sensor), so widespread use in a harsh environment is not cost effective. Therefore, it served as the control to which another PPG watch sensor would be compared. The second sensor, a lower cost (i.e. ~$250), commodity sensor, was the Apple Watch. This device was chosen because it outperformed several other commodity heart rate watches in both accuracy and precision (El-Amrawy et al., 2015). Several potential issues could cause these wearable sensors to be ineffective at stress monitoring. For example, improper placement on the body during tasks and large arm motions could cause the position of the PPG sensor relative to the skin to move, causing interference with the measurement of reflected light, resulting in skewed or null data points. In this study, participant's heart rate was recorded by both sensors throughout the AR assembly tasks. The data were then compared and analyzed for significant changes in heart rate which could indicate stress due to the AR instructions or the current assembly step. By comparing the heart rate results of the commodity sensor to that of the medical grade sensor, the viability of this technology as a method of evaluating stress on participants using AR work instructions could be determined.

BACKGROUND

The background section will present research and the state-of-the-art in a few key areas. First, current and past projects that studied heart rate as a physiological marker of stress will be outlined. Second, research projects focusing on the development of AR systems for assembly or maintenance tasks will be discussed. Lastly, current AR evaluation techniques will be described and the potential use of heart rate sensing watches for this purpose will be proposed.

Heart Rate as a Physiological Marker of Stress

Determining which physiological signals can be used to accurately determine levels of stress is an ongoing topic of research in the field of physiological sensors. One such study specifically analyzed physiological reactions while driving (Healey & Picard, 2005). By comparing driver stress levels to data acquired from an ECG, electromyogram (EMG), galvanic skin response (GSR) sensor, and respiration monitor, the researchers were able to determine that heart rate monitoring and GSR were the most accurate of the methods tested for determining stress. However, this research was limited to the study of devices with tethered connections. Today's wearable physiological sensors could make the idea of real time driver stress analysis a reality by providing a comfortable, wireless way to acquire physiological data. Other recent studies have also tested for heart rate monitoring as a method for quantifying stress with success. In one study, 43 participants performed office tasks of low and high mental load involving arithmetic and clicking tasks using a computer while equipped with an ECG (Taelman, Vandeput, Vlemincx, Spaepen, & Van Huffel, 2011). By examining the changes in heart rate variability (HRV), the researchers were able to differentiate between the mental load and rest conditions.

To date, many user studies have been conducted, in controlled laboratory conditions, to study physiological stress monitoring techniques. Fewer researchers have studied the accuracy of these sensing methods during realistic tasks, or even in the presence of physical activity. A 2010 paper by Sun et al. is one of the few which confronted this problem. In the study, participants were equipped with ECG as well as GSR devices and an accelerometer (Sun, Kuo, Cheng, & Buthpitiya, 2010). The researchers found that by using a combination of these three sensors, they could accurately detect cognitive stress despite the varying levels of physical stress.

Another study that demonstrates the effectiveness of PPG for stress detecting applications was reported by Lyu et al. (2015). Lyu et al. created a new method of quantifying stress called the stress-induced vascular response index (sVRI). It uses data from a finger PPG sensor as opposed to ECG. The method was verified by a user study where participants were given cognitive tests of three different difficulty levels. The results showed that this PPG based method was equally as accurate as other popular physiological stress measurements. Although this paper represented a large step in stress measurement using PPG sensors, a less obtrusive sensing device must be employed for tasks requiring more free movement. However, the immaturity of wrist worn PPG sensor technology brings into question the accuracy and precision of these somewhat new devices, especially for those not approved for medical use. A recent study answered this question by comparing the step tracking and heart rate monitoring capabilities of 17 different commodity wearable devices. User's heart rates were recorded by the wearable devices and compared to the reading of a clinical pulse oximeter. The results showed that an Apple Watch outperformed other devices such as the Samsung Gear and the Motorola Moto 360 in both heart rate accuracy and precision (El-Amrawy et al., 2015). These results supported the choice of an Apple Watch for this work.

Evaluation Techniques for Augmented Reality

This paper focuses on the potential of wearable heart rate sensors to help improve the evaluation and delivery of AR assembly instructions. Several studies have evaluated the effectiveness of AR in industry. Specifically, Gavish, Gutiérrez, & Webel compared the effectiveness of AR training instructions with traditional training using video (Gavish et al., 2015). They found that technicians trained using AR performed fewer errors in an industrial maintenance and assembly task than the control. Henderson & Feiner developed and evaluated a head mounted AR maintenance aid for the turret of an Armored Personnel Carrier (Henderson & Feiner, 2009). The study used timed tasks and self-reported measures to conclude that an AR system allowed military maintenance personnel to find task locations more quickly than with traditional LCD monitor based instructions. However, it was also discovered that users found the LCD monitor easier to use than AR technology.

A previous study conducted at Iowa State University's Virtual Reality Applications Center (VRAC) analyzed the use of motion tracking including head movement and orientation to help determine how often users referred to the tablet or computer monitor to verify work instructions (Richardson et al., 2014). The study concluded that participants looked at the tablet less frequently when using AR instructions than when traditional model based instructions were used.

Studies like these have made it clear that AR work instructions are advantageous. By exploring non-traditional methods of continuous user monitoring, AR technology can be analyzed and refined in a more efficient manner. Specifically, heart rate monitoring PPG sensors have been proven to detect stress in many different applications. By employing wearable wrist PPG monitors to users performing AR assembly tasks, it may be possible to pinpoint steps inducing higher stress on a user. This information could then be used to improve the AR work instructions or mitigate the stress of the worker before safety or production quality issues arise. In order to do this, the ability of the sensors to detect changes in heart rate under these conditions must be verified.

USER TESTING

A user study was conducted in order to test the effectiveness of using wrist worn sensors as a method to evaluate AR work instructions. During the study, participants assembled a mock wing in a simulated industrial work cell. The following sections describe the technology and procedure used for this study.

Augmented Reality System Development

The interface used to present the AR work instructions for this study was created using ARmaker. ARmaker is an in-house augmented reality application which is built off of two software libraries: OpenSceneGraph and ARToolKit. ARmaker allows developers to create an AR application which renders computer graphics in real time over the video scene. The graphics were accurately positioned on the scene using data from a four-camera Vicon infrared tracking system. Stationary objects, such as part bins and tables, as well as the mobile tablet stand and helmet, were equipped with reflective marker balls which were tracked using four infrared cameras positioned around the work cell. The marker based tracking system allowed the AR application to properly orient virtual objects over real objects on the screen as shown in Figure 1.

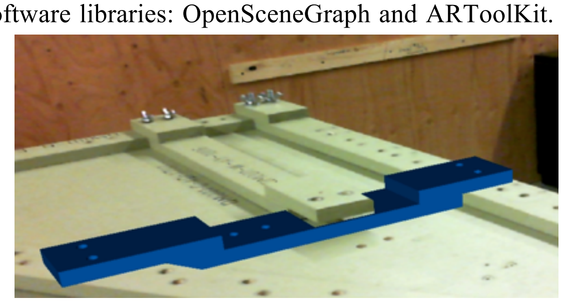

Figure 1. AR work instruction with occlusion.

The user interface used animated 3D part models to demonstrate the assembly tasks. Occlusion via contours was used to help users understand if parts belong in front of or behind other real objects during complex steps. It was important that occlusion was included in order to simulate depth perception and to not confuse the user. Figure 1 shows an example of a step using occlusion to indicate the proper placement of two overlapping spars.

Hardware

During the user study, the AR work instructions were delivered using an 11-inch Dell Venue Pro tablet (1.6 GHz Intel Core i5 processor) equipped with a Microsoft LifeCam HD-6000 for Notebooks (30 fps). The tablet and web camera were mounted on a mobile stand with a pivoting arm. Wheels attached to the bottom of the stand allowed the participant to move the tablet freely around the room with little effort. The stand, as well as the various tables throughout the room, were tracked using the Vicon tracking system. The observer station consisted of an iMac desktop PC equipped with Windows 8 and a MacBook Air laptop. The iMac was used to make notes during the task. The MacBook Air was positioned in the corner of the room and was equipped with an application to record the movements of all the tracked objects during the study.

Two different PPG sensors were worn by the participant (one on each wrist) throughout the study to record heart rate data. The first device was an Empatica E4 which is a high end device aimed at providing the most accurate data for health and diagnostic applications (Garbarino, Lai, Bender, Picard, & Tognetti, 2015). The Empatica E4 records HR data at a rate of 1 Hz. The second PPG sensor was an Apple Watch. The Apple Watch represents the top of the line for lower cost commodity sensors typically used to record heart rate for physical fitness and daily activity logging rather than medical monitoring (El-Amrawy et al., 2015). The Apple Watch was interfaced with an Apple iPhone 6 plus via blue tooth and is capable of recording HR data at 0.2 Hz and is capable of measuring 3-axis acceleration. While both sensors are capable of capturing more data, only the heart rate data was used for this study. A more detailed description of each device is shown in Table 1, below.

Table 1. Comparison of Empatica E4 and Apple Watch.

	Empatica E4	**Apple Watch**
Cost	$1,690	$299
Sensors	PPG GSR* Acceleration* Skin temperature*	PPG Acceleration*
Sample Rate	1 Hz	0.2 Hz
Medical Regulatory Compliances	CE Medical 93/ 42/EEC Directive, class 2A	None
Battery Life (while streaming)	36 hours (20 hours)	18 hours (6.5 hours)
* Sensor functionality not used in this study		

Assembly Task

In order to make the mock wing assembly as realistic as possible, the research team worked closely with employees of The Boeing Company while conceptualizing the assembly tasks and fabricating the materials. As previously stated, the participants of this study were asked to assemble the mock wing using wooden spars, nuts, bolts, washers and wires. The parts were assembled on a triangular table in the center of the work cell as shown in Figure 2. Fasteners were gathered from a shelving unit containing 20 different fastener bins. The bins were labeled with part numbers corresponding to the fasteners within them. Wooden spars were collected from a large parts table. Each spar was labeled with a unique part number.

Figure 2. Completed wing assembly.

The assembly task consisted of 46 steps delivered using AR work instructions. At the beginning of each step, a series of navigation windows called "gates" appeared. These gates led the user to the area of the work cell needed to complete that particular step. After the point of interest was located, an animated 3D object or part model would appear to indicate what task should be performed. These depended on the nature of the step: 1) choosing a part, 2) orienting a part on the table or 3) installing fasteners or hardware. In the case of selecting correct fasteners, the proper parts bin was surrounded by a green box and contained a number indicating the quantity needed as shown in Figure 3. In the case of demonstrating how to orient a part or install hardware, an animation showing models of all parts and fasteners would appear. The animation displayed the necessary paths of the parts beginning with an exploded view and ending with the fully assembled model. This animation repeated until the user advanced to the next step.

Users advanced through the steps using a check mark box on the lower right corner of the tablet screen. The user could also navigate backwards to check their work by using a step navigation menu located on the right side of the interface. Users had the option of displaying or hiding this menu throughout the study.

Procedure

Each study began with a participant being welcomed and asked to read an informed consent document. After agreeing to and signing the document, participants were immediately equipped with two wrist worn PPG sensors: the Apple Watch and the Empatica E4, for continuous heart rate monitoring throughout the study. (The hand on which each sensor was worn was determined at random for each participant.) Next, each participant was asked to complete a pre-survey, which was used to acquire demographic data including age, gender, education, assembly confidence, and technology usage. Participants were informed that they may skip any questions they preferred not to answer. During this time, participants were seated and in a mostly sedentary state. Because of this, heart rate data recorded during this time was used to determine the participant's resting heart rate.

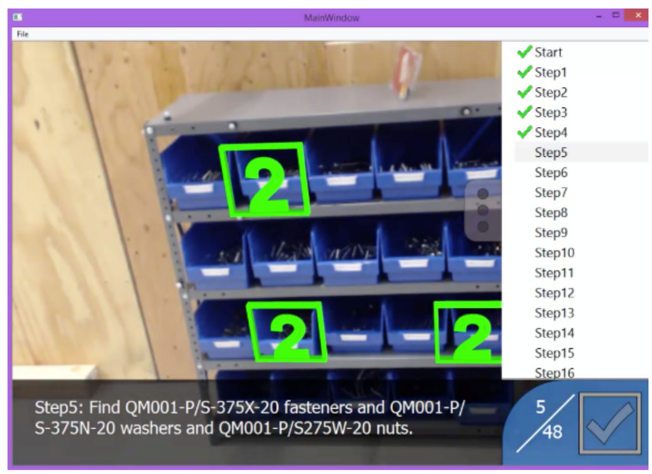

Figure 3. AR part selection.

After the questionnaire, the participant was instructed to complete a short practice assembly using wooden spars and fasteners similar to those used in the mock wing assembly. First, the observer instructed the participant on how to use the AR interface and the function of various objects throughout the room. The participant was also instructed to ask clarifying questions during this phase of the study in order to acclimate themselves to the interface and the work cell. The participant was made aware that there was no time limit for the practice trial. During the practice trial, as well as the wing assembly phases, the participant wore a motion tracked hard hat, and acoustic guitar music was played softly in the background.

Upon completion of the practice assembly, the user was instructed to assemble the mock wing for the first time. The participant was again instructed on how to use the AR interface as well as the tables and parts throughout the room. However, this time, the participant was notified that no questions regarding the user interface or assembly tasks would be answered during the trial and that they would be evaluated on both speed and accuracy of assembly. They were also told that a time limit of 45 minutes would be imposed. After the assembly was complete, the observer graded and disassembled the mock wing in preparation for the second trial. Assembly accuracy was graded based on three criteria:

the necessary parts were present, the parts were positioned in the correct location or orientation, and there were no extraneous parts present. During a short intermission between the first and second wing assembly trials, participants were given a three minute paper folding test (Ekstrom, French, Harman, & Dermen, 1976). The purpose of the paper folding test was to evaluate the participant's spatial thinking capability. The paper folding test also allotted the observer time to grade and disassemble the mock wing.

A second trial of the identical mock wing assembly was performed in the same manner as the first trial described. This assembly was graded and disassembled while the participant completed a post-study questionnaire. This questionnaire consisted of several self-reported rankings as well as free response questions regarding the effectiveness of the AR interface and the participant's experience. After turning off and removing the two wrist worn sensors, the participant was thanked and compensated $20 for their time.

RESULTS AND DISCUSSION

The results reported in this section were calculated using data from ten participants. A small number of participants were used in this initial study in order to determine if the use of wearable heart rate monitors is a feasible method of AR evaluation. All participants were engineering students between the ages of 18 and 22. Of the group, two participants were female, and eight were male. In the subsections below, the data recorded by the Empatica E4 and the Apple Watch will be compared. First a correlation analysis was performed, followed by a comparison of heart rates for three different "categories" of steps, and lastly, heart rates were compared for steps of different difficulty levels.

Device Correlation

In order to determine if a linear correlation was present between the heart rate data collected by the Apple Watch and the Empatica E4, the data was consolidated into average heart rate per assembly step for each device. This was carried out by calculating the mean of all heart rate measurements collected during an assembly step as determined by the time stamps from the AR application log files. It should be noted that fewer data points were used to calculate mean heart rate values for the Apple Watch because of its lower sample rate and occasional null data points likely caused by jostling of the wrist sensor. This process was repeated for all 46 steps for each trial of the wing assembly. A linear correlation analysis was performed to determine if the reading for the two PPG watch sensors were similar. If both of the sensors consistently recorded accurate heart rate data, one could expect to see an obvious linear correlation with a value of r ≈ 1. The scatter plot in Figure 4 shows little linear correlation between the Apple Watch and the Empatica E4 heart rate values, this is supported by the Pearson correlation, which indicated only a moderate correlation of $r=0.302$, $n=607$, $p<.0005$. This likely indicates that the heart rate data acquired by one of the sensors is more accurate than the other because the values did not have a strong, positive, linear correlation.

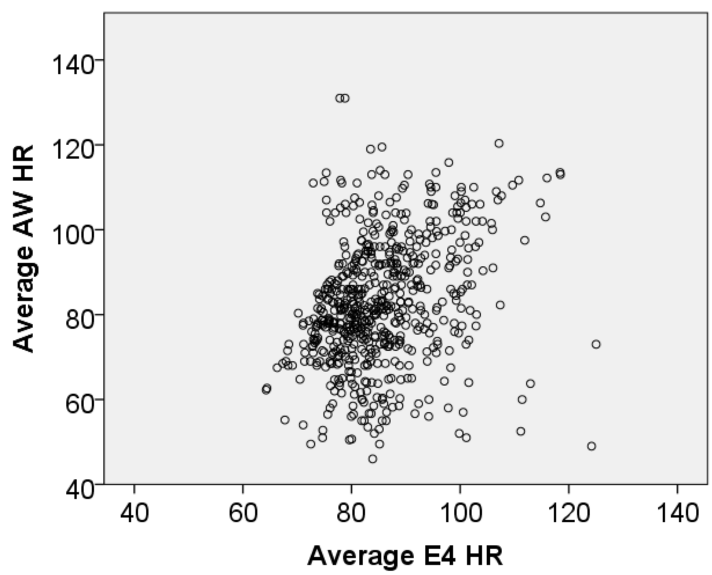

Figure 4. Graph of correlation between Apple Watch and E4 data.

Task Type

In next step of the analysis, the average step heart rates calculated in the previous subsection were used to determine if a significant difference could be found between steps of different types. In order to determine if the mean heart rate was different between steps of different types, the steps were first categorized. Three categories of steps used were:

1) **Picking steps:** participants located and picked up the part(s). An example of a picking step can be seen in Figure 3.
2) **Placing steps:** participants placed a spar onto the wing assembly table in the correct location and orientation as indicated by the AR interface (Figure 1).
3) **Assembling steps:** participants secured parts to the table or to other parts on the table using fasteners or wires as is indicated by the work instruction example shown in Figure 5 and Figure 6.

It was hypothesized that the different types of steps may induce different levels of cognitive load on the user causing a difference in heart rate between these three step types. However, the results of a one-way ANOVA showed that there was no significant difference in mean heart rate between picking, placing and assembling steps for either device. Although, there was no significant difference, an interesting trend did appear indicating that some participants may have had lower mean heart rates during the placing steps than during the picking and assembling tasks for both devices. However, it is overshadowed by the large standard deviations from the Apple Watch compared to those of the Empatica E4. These large standard deviations in Beats Per Minute (BPM) were possibly a result of the frequent hand motion necessary for the assembly of parts. The Apple Watch standard deviations, especially during assembling steps (15.30 BPM), are particularly large when compared to those of the Empatica E4 (9.54 BPM). This may be caused by the lack of motion artifact mitigation by the Apple Watch and the smaller sample size due to its low sample rate. The heart rate values recorded by the Apple Watch (82.77 ± 14.95 BPM) also tended to be lower than those recorded by the Empatica E4 (86.30 ± 9.69 BPM). The mean values, as well as the standard deviations of all three step types can be seen in Table 2.

Table 2. Mean heart rates by step type.

Device	Picking Steps	Placing Steps	Assembling Steps	All Steps
Empatica E4	86.19 ± 9.73 BPM	84.34 ± 9.78 BPM	87.13 ± 9.54 BPM	86.30 ± 9.69 BPM
Apple Watch	82.78 ± 14.95 BPM	80.83 ± 13.64 BPM	83.25 ± 15.30 BPM	82.77 ± 14.95 BPM

*mean ± 1 standard deviation

Task Difficulty

For this portion of the data analysis, assembly steps were divided into two categories of difficulty: simple and complex. The majority of the steps were determined to be simple, while five of the 46 steps (step number 20, 25, 42, 44 and 46) were assigned to the more complex category. These steps were deemed "complex" because previous iterations of the study showed that participants made more errors during these steps (Richardson et al., 2014). Step 20 consisted of a placing step during which participants had to twist a spar into position. An image of the AR instructions can be seen in Figure 1. This step can be confusing because the spar fit very tightly, and only the precise twisting motion indicated in the AR instructions would allow the spar to fit. It was also very easy to insert the spar backwards causing the holes to misalign. Another more difficult step, step 25, instructed the participant to install a spar using fasteners and a metal washer (Figure 5). This was very similar to many other steps in the assembly, however, this one instructed the user to place the washer underneath the spar instead of on top of it. This was found to cause confusion and many errors for previous participants. The last three steps in this category, steps 42, 44, and 46, were very similar, and can be grouped together. These steps instructed the user to insert wire "stringers" through holes in

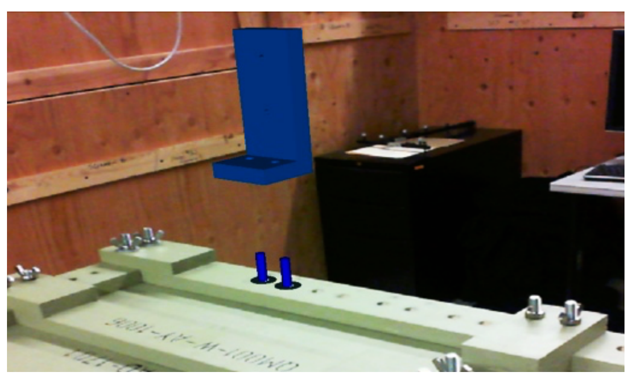

Figure 5. Assembling step 25 with washer placement beneath spar.

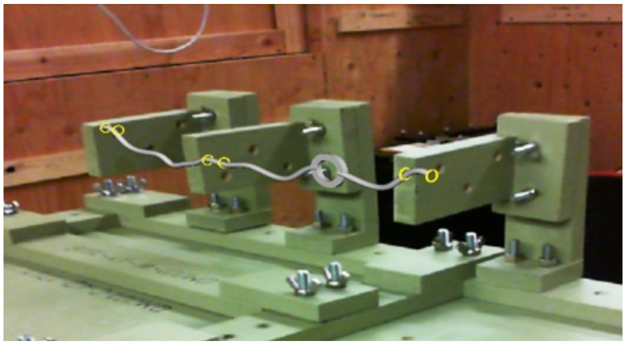

Figure 6. Step 42 misalignment with spar holes.

three different spars (Figure 6). During this step, participants had to be careful to string the wires through the correct holes and align the metal washers appropriately. Previous participants often made mistakes during these steps because the AR interface did not always perfectly align the wire with the correct hole, forcing the user to either guess where the wire belonged, or reposition the tablet to attain another view point.

In order to make a comparison of heart rate between simple and complex steps, two independent samples t-tests were performed. The results of these t-tests showed that the mean step heart rates for the Empatica E4 were significantly higher for the complex steps (M=89.56, SD=11.31) than for the simple (M=85.90, SD=9.41) steps; (t(104.584)=2.956, p=0.004). However, the same conclusion could not be drawn for the Apple Watch heart rate data. There was no significant difference in the mean step heart rate detected by the Apple Watch between the simple (M=82.73, SD=15.00) and complex steps (M=83.10, SD=14.72). A plot of means showing this data can be found in Figure 7. The error bars in this plot indicate one standard deviation. The inability of the Apple Watch to discern differences in heart rate between these tasks indicates that it may not be a suitable sensor for this application. This could be caused by a number of factors including the sensor's low sample rate or, potentially, an inability to cope with large amounts of hand motion and perspiration that can interfere with the measurement of light reflectance by the PPG sensor.

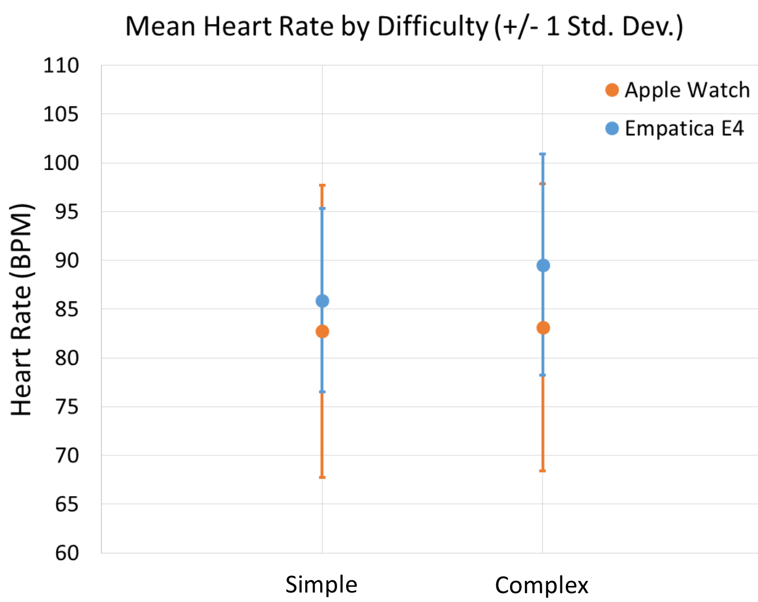

Figure 7. Mean heart rates by step difficulty with standard deviations.

The results have clearly indicated that the commodity Apple Watch sensor is not accurate enough to detect heart rate changes which may indicate stress or cognitive load during AR assembly tasks. In order for this device to be a viable sensor for military AR applications, the technology would require improvements such as a more frequent sample rate and the removal of motion artifacts which may interfere with the heart rate reading. However, with the continued refinement of commodity sensing devices, in the future they may be accurate enough to produce conclusive results like the Empatica E4. Alternately, the Empatica E4 showed some ability to differentiate between steps using heart rate data. A continuation of this study, currently on-going, will attempt to verify this data with a larger sample size. By validating and developing this method of AR evaluation, these wearable heart rate sensors can be a valuable and economical tool for today's warfighter.

CONCLUSIONS AND FUTURE WORK

Based on the user study data analysis, heart rate data recorded by the Apple Watch was not found to be a viable option for evaluating AR work instructions for military applications. The Apple Watch was not able to detect differences in heart rate among steps of different type or difficulty. The Empatica E4 device, on the other hand produced heart rate data which was significantly different between steps of low and high difficulty. Some other trends in the Apple Watch heart rate data also lent themselves to the conclusion that the Apple Watch data could have some inaccuracies. For example, the Apple Watch data had a much larger standard deviation than the Empatica E4 data which could have been caused by noise or by the smaller sample size of the data due to the lower sample rate. Another poor indicator for the Apple Watch was the presence of gaps in data collection. These gaps were most likely a result of participant's movement, specifically, their hand movement, while assembling parts during the study. The same dislodgement is possible for the Empatica E4, however, the Empatica E4 device incorporates motion artifact filtering to minimize this effect, making it a more suitable device for detecting heart rate during assembly tasks. Overall, a medical grade sensor, like the Empatica E4 is a more viable candidate sensor for this application as opposed to lower-cost commodity sensing devices, as evidenced by the results of the Apple Watch. The use of a medical grade sensor, like the Empatica E4 would be more beneficial for evaluating the AR work instructions and ultimately helping today's warfighters do their jobs more accurately and efficiently.

Now that there is evidence that wrist worn heart rate sensors have the potential to pinpoint stress inducing steps within AR work instructions, this technology can be advanced in order to help improve the efficiency of AR for military maintenance personnel. For example, a similar wearable heart rate sensing method could be used to identify difficult or stressful steps during training situations. If implemented in real time, this could allow the training process to be more efficient and military maintenance staff to become more accurate and confident when performing assembly tasks. Future work on this project includes expanding the study to 35 to 40 users and testing the heart rate sensors in a real factory environment instead of the simulated work cell. The improved and expanded user study would provide verification of the results found in this study and allow the researchers to discover any problems which might arise in an actual industrial environment. Additionally, an expanded study could be used to analyze the effect of wearing the sensor on the dominant versus the non-dominant arm. Other possibilities for the progression of this research include testing other known physiological markers for stress, such as GSR, and identifying the specific heart rate trends which can indicate stress for AR assembly instruction users.

REFERENCES

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*. Retrieved from http://iopscience.iop.org/article/10.1088/0967-3334/28/3/R01/meta

Azuma, R. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/pres.1997.6.4.355

Bidwell, J., Khuwatsamrit, T., Askew, B., Ehrenberg, J. A., & Helmers, S. (2015). Seizure reporting technologies for epilepsy treatment: A review of clinical information needs and supporting technologies. http://doi.org/10.1016/j.seizure.2015.09.006

Boff, K., Kaufman, L., & Thomas, J. (1986). Handbook of perception and human performance. Retrieved from https://www.researchgate.net/profile/George_Sperling/publication/224927334_Handbook_of_human_perception_and_performance/links/54b2aae20cf220c63cd272e9.pdf

Dünser, A., Grasset, R., & Billinghurst, M. (2008). *A survey of evaluation techniques used in augmented reality studies*. Retrieved from https://www.researchgate.net/profile/Mark_Billinghurst/publication/216867631_A_Survey_of_Evaluation_Techniques_Used_in_Augmented_Reality_Studies/links/09e4150782be47b86d000000.pdf

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Manual for kit of factor referenced cognitive tests*. Retrieved from http://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf

El-Amrawy, F., Nounou, M. I., McGrath, M., Scanaill, C., McGrath, M., Scanaill, C., … Patel, M. (2015). Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? *Healthcare Informatics Research*, *21*(4), 315. http://doi.org/10.4258/hir.2015.21.4.315

Garbarino, M., Lai, M., Bender, D., Picard, R. W., & Tognetti, S. (2015). Empatica E3 - A wearable wireless multisensor device for real-time computerized biofeedback and data acquisition. *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", MOBIHEALTH 2014*, 39–42. http://doi.org/10.1109/MOBIHEALTH.2014.7015904

Gavish, N., Gutiérrez, T., & Webel, S. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning*. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10494820.2013.815221

Healey, J., & Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems,*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1438384

Henderson, S., & Feiner, S. (2009). Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *Mixed and Augmented Reality, 2009.* Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5336486

Hymel, K. M. (2015). BATMAN ' S friend , BATDOK , uses technology to save lives. Retrieved from http://www.af.mil/News/ArticleDisplay/tabid/223/Article/626739/batmans-friend-batdok-uses-technology-to-save-lives.aspx

IDC. (2015). IDC Forecasts Worldwide Shipments of Wearables to Surpass 200 Million in 2019, Driven by Strong Smartwatch Growth. Retrieved May 12, 2016, from https://www.idc.com/getdoc.jsp?containerId=prUS40846515

Lyu, Y., Luo, X., Zhou, J., Yu, C., Miao, C., Wang, T., … Kameyama, K. (2015). Measuring Photoplethysmogram-Based Stress-Induced Vascular Response Index to Assess Cognitive Load and Stress. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (pp. 857–866). http://doi.org/10.1145/2702123.2702399

Muaremi, A., Arnrich, B., & Tröster, G. (2013). Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep. *BioNanoScience*, *3*(2), 172–183. http://doi.org/10.1007/s12668-013-0089-2

Rawassizadeh, R., Price, B. A., & Petre, M. (2015). Wearables: Has the Age of Smartwatches Finally Arrived? *COMMUNICATIONS OF THE ACM*, *58*(45). http://doi.org/10.1145/2629633

Richardson, T., Gilbert, S., Holub, J., Macallister, A., Radkowski, R., Davies, P., & Terry, S. (2014). Fusing Self-Reported and Sensor Data from Mixed-Reality Training. *I/ITSEC*, (14158), 1–12. Retrieved from http://www.frederickt.com/pubs/iitsec2014.pdf

Sun, F., Kuo, C., Cheng, H., & Buthpitiya, S. (2010). Activity-aware mental stress detection using physiological sensors. *Mobile Computing,*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-29336-8_12

Taelman, J., Vandeput, S., Vlemincx, E., Spaepen, A., & Van Huffel, S. (2011). Instantaneous changes in heart rate regulation due to mental load in simulated office work. *European Journal of Applied Physiology*, *111*(7), 1497–1505. http://doi.org/10.1007/s00421-010-1776-0

U.S. Department of Defense. (2015). DoD Announces Award of New Flexible Hybrid Electronics Manufacturing Innovation Hub in Silicon Valley. Retrieved October 5, 2016, from http://www.defense.gov/News/News-Releases/News-Release-View/Article/615132/dod-announces-award-of-new-flexible-hybrid-electronics-manufacturing-innovation

World Healthcare Organization. (2003). Medical device Regulations Global Overview and Guiding Principles. *Journal of Healthcare Information Management : JHIM*, (3), 36–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21530908

Yoo, K., & Lee, W. (2011). Mental stress assessment based on pulse photoplethysmography. *Consumer Electronics (ISCE), 2011 IEEE 15th*. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5973841