Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation

Xianjun Sam Zheng
Siemens Corp., Corporate Technology
755 College Road East
Princeton, NJ, USA
sam.zheng@siemens.com

Cedric Foucault
Siemens Corp., Corporate Technology
755 College Road East
Princeton, NJ, USA
cedric.foucault@siemens.com

Patrik Matos da Silva
Siemens Corp., Corporate Technology
755 College Road East
Princeton, NJ, USA
patrik.matos@siemens.com

Siddharth Dasari
Siemens Corp., Corporate Technology
755 College Road East
Princeton, NJ, USA
siddharth.dasari@siemens.com

Tao Yang
Siemens Corp., Corporate Technology
755 College Road East
Princeton, NJ, USA
tao.yang@siemens.com

Stuart Goose
Siemens Corp., Corp. Tech., Technology to Business
1995 University Ave Berkeley, CA, USA
stuart.goose@siemens.com

ABSTRACT
Exciting developments in eye-wearable technology and its potential industrial applications warrant a thorough understanding of its advantages and drawbacks through empirical evidence. We conducted an experiment to investigate what characteristics of eye-wearable technology impact user performance in machine maintenance, which included a representative set of car maintenance tasks involving Locate, Manipulate, and Compare actions. Participants were asked to follow instructions displayed on one of four technologies: a peripheral eye-wearable display, a central eye-wearable display, a tablet, or a paper manual. We found a significant effect of display position: the peripheral eye-wearable display resulted in longer completion time than the central display; but no effect for hands-free operation. The technology effects were also modulated by different Tasks and Action types. We discuss the human factors implications for designing more effective eye-wearable technology, including display position, issues of monocular display, and how the physical proximity of the technology affects users' reliance level.

Author Keywords
eye-wearable technology; machine maintenance; action type; task guidance; wearable computing; smartglasses.

ACM Classification Keywords
H.5.m. Information interfaces & presentation: Misc.

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3145-6/15/04…$15.00.
http://dx.doi.org/10.1145/2702123.2702305

INTRODUCTION
Recent advancement of eye-wearable technology, such as Google Glass, has generated a lot of excitement regarding its potential applications, especially in industrial settings. The value of the smartglasses market is expected to reach nearly $6 billion by 2020 [26], and Gartner predicts that smartglasses can yield savings of $1 billion annually for industrial field service [6]. However, as the end-user reviews for such technologies have been mixed, we need to gain an understanding of its fundamental characteristics and their impact on human performance through empirical evidence.

We define eye-wearable technology as any technology having a digital display that is worn close to the eye and does not fully obstruct the real world view; this is also referred as smartglasses, or digital eye glasses. Thirty years ago, Mann [14] designed the first kind of this technology, a multimedia wearable computer that allowed him to walk around while keeping an eye on the screen. Since then technical breakthroughs have been achieved in miniaturization, power consumption, and networking, and we are now reaching the productization stage, with manufacturers such as Google, Epson, and Vuzix, introducing new offerings to the market.

Researchers have conjectured that by making information constantly accessible at a glance without monopolizing the user's attention, and by making the interaction hands-free, eye-wearables can support the user's primary task better than any other technology. However, empirical evidence is lacking to support the benefits claimed. The technological factors of eye-wearables have not been clearly teased out, meaning that two eye-wearable systems may have very different results. Those factors include the eye-wearable display's position, opacity, and field-of-view coverage. Furthermore, applications in industrial settings, and in particular machine maintenance, have not yet been thoroughly investigated.

As of June 2014, the machine maintenance industry has $35 billion revenue and employs 262,447 people [12]. Despite being a sizable industry, machine maintenance is still relatively low-tech and relies heavily on paper-based systems. Each asset has several paper files containing the asset's detailed description, maintenance procedures, and past maintenance records, etc. Conventional computer technology has had a limited impact on this industry largely because of the uncontrolled environment involved in the field and the technicians' limited interaction capability while performing their tasks. Wearable computers are designed to work in those contexts by moving computing onto the body, and thus have the potential to revolutionize this industry. In this paper, we focus on one application: using eye-wearable technology to display task instructions in preventative car maintenance.

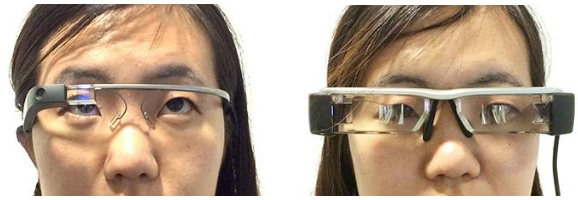
As of 2010, there are more than 1 billion cars in operation worldwide [36], and each one needs periodic maintenance. Because cars are easily accessible, experimenters can use them in the setting they want to study (indoors, outdoors, etc.). The popularity of do-it-yourself car maintenance is evidence that there are many potential candidates for such experiments.

We devised an experimental method for studying machine maintenance that is easily replicable and also provides a realistic task environment. We use the car as a convenient and ecologically valid research vehicle. Our study seeks to better understand the implications of the attributes that are essential to eye-wearable technology from a human factors perspective and highlight the resulting trade-offs.

RELATED WORK

Design and Applications of Eye-Wearable Technology

Despite small variations in eye-wearable implementation [14, 22, 27], there are two main approaches for design: one is to put a see-through display at the center of the user's line-of-sight, and the other one is to position a display at the eye's periphery (see Figure 1 for an example). In the first approach, which we refer as Eye-Wear Central, users can see both the real world and the instruction in their central vision. However, switching between the two information channels requires ocular accommodation since the two object domains do not converge at the same distance, an eye movement called *vergence*. Moreover, it is difficult to display rich visual information without obstructing the user's view of the real world. In the second approach, Eye-Wear Peripheral, a smaller display is typically placed above the line-of-sight. The information can be viewed by literally looking up as with a car's rear-view mirror. One hypothesized benefit over a direct line-of-sight display is that users can better focus on the real-world task, since the information does not obstruct their central vision.



(a)Google Glass (b)Epson Moverio

Figure 1. A user is wearing the device used in (a) Eyewear-Peripheral & (b) Eyewear-Central conditions.

This, however, requires users to move the eye up and to the side to look at the display, an eye movement called *extorsion* [9].

Applications of eye-wearable technology include contextual reminders and agents, remote communication, and augmented reality (AR) applications, such as a museum guide and industrial maintenance [4,7,10,15,22,28]. A number of research groups developed AR applications for machine maintenance, field service, and manufacturing [3, 5,7,8,30,31], and they suggested benefits such as reducing the number of personnel and time required to perform tasks. Other industrial applications included stored procedures retrieval, synchronous/asynchronous collaboration, and context-aware platforms with a proactive assistant [10, 25]. In particular Siewiorek et al. [24] developed a wearable system for train maintenance providing the ability to retrieve information and navigate through it to follow particular working procedures. However, rigorous evaluations were lacking for these systems [1].

Studies and Empirical Work

Most empirical work that involves eye-wearable technology based task guidance has focused on comparing a wearable system with the status quo of the domain under study. Siegel & Bauer [23] conducted a field study comparing a wearable system with a paper technical order on two aircraft maintenance tasks. Although avionics showed enthusiasm for the wearable system, they took on average 50% more time to perform the tasks with the wearable system than with the paper technical order. Ockerman & Pritchett [17] conducted a study on the preflight inspection of an aircraft in which they compared three inspection methods: the regular by-memory method and two methods using a HMD-based wearable system, one with text only, the other one with text and pictures. They found no statistically significant effect on fault detection rate, but found that the pilots using any of the two wearable computers forgot to check items that were not mentioned on the computer significantly more often than the ones doing the inspection by memory.

One limitation of those studies is that the wearable systems they evaluated were early prototypes, often already complex to use, and the participants were using for the first time, whereas they had professional knowledge of the status-quo method. Furthermore, they did not tease out the factors differentiating the system that they evaluated from the control method that may have effects on task performance. For example, Siegel & Bauer [23] used a complex system that integrated a HMD, a belt-worn pen-based touchscreen, and a belt-worn dial. It is difficult to conclude that their results can be transferred to other eye-wearable technology based systems, even with a similar use case.

Contrary to the studies mentioned above, Weaver et al. [33] did find that participants were significantly faster when using the HMD based method than when using any of the other three methods (paper graphic, paper text, and audio text) in a warehouse picking task. The HMD method led to significantly less errors, higher usability rankings, and lower overall task load than the text-based paper method. This study was better controlled: they used an order picking environment specifically built for the experiment; they leveraged the Wizard of Oz technique to control the effects of input modality. However, this experiment focused on one specific task, picking items on a shelf, and uses one specific representation, which was optimized for the layout of a specific indoor environment. It is unclear whether the observed effects would translate to preventive maintenance tasks. By contrast, we are studying car maintenance tasks, which can be performed indoors or outdoors, and involves a variety of tasks and tools, placing higher physical and cognitive demand on the worker.

None of the studies mentioned above included a mobile technology condition, so the effects of the eye-wearable technology could also be attributed to the mobility factor. In addition, they included only one eye-wearable condition, however different eye-wearable form factors such as position of the display could also lead to different results. Therefore, we included a mobile condition and two eye-wearable conditions that we found most representative in our experiment.

METHOD

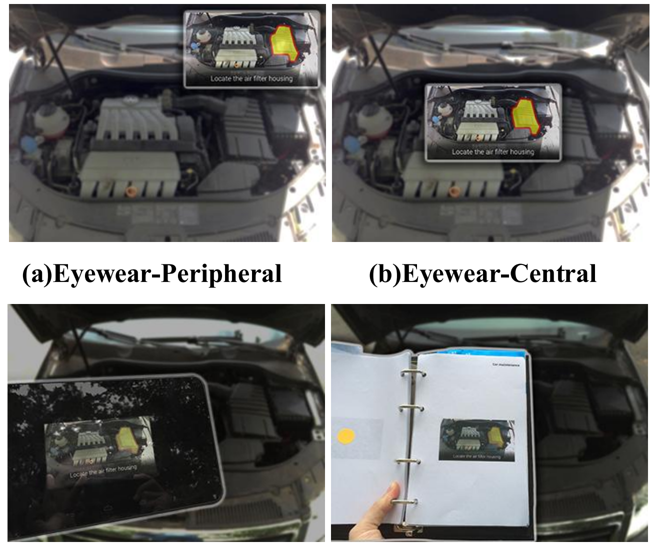
Our goal is to investigate how different eye-wearable technology attributes affect user performance on realistic maintenance tasks. We studied three independent variables: Technology, Task, and Action Type. We conducted a experiment involving a varied set of car maintenance tasks and four experimental conditions that exhibit eye-wearable and non eye-wearable factors. In all four conditions, participants were provided step-by-step instructions that were identical for the same task.

Technology Conditions

Figure 2 shows a first-person view of the information displayed by each of the four technology conditions.

Eyewear-Peripheral

In this condition, the display is eye-worn and the current instruction is displayed above the participant's line-of-sight. To navigate the instructions, participants had to use voice



(a)Eyewear-Peripheral **(b)Eyewear-Central**

(c)Tablet **(d)Paper**

Figure 2. A first-person view of the information displayed by each of the four technology conditions.

commands: "Next" to go one step further, and "Previous" to go one step back. We used the Wizard of Oz technique to interpret the participant's command: a human wizard listened to the participant and initiated the appropriate response remotely.

The display system we used to operationalize this condition is Google Glass (Explorer Edition) (Tech Specs: 60g, 560 × 530 mm headset, monocular eyeglass, prism projector, Android 4.4). It contains a monocular, semi-transparent, eye-wearable display, positioned in the upper-right hand corner of the right eye, and it is lightweight (See Figure1(a)).

Eyewear-Central

In this condition, the display is eye-worn and the current instruction is displayed directly in front of the participant's line-of-sight. The same navigation technique is used as with the Eyewear-Peripheral condition.

The display system we used is the Epson Moverio BT-200 Smart Glasses(Tech Specs: 88g, 170 × 185 × 32 mm headset, binocular glasses, see-through LCD, Android 4.0). It contains a binocular, semi-transparent, eye-wearable display, it is worn as regular glasses and it is medium-weight. (See Figure1(b)).

Tablet

In this condition, the current instruction is displayed on a mobile tablet computer. The same navigation technique is used as with the Eyewear-Peripheral condition. Compared to the previous two conditions, the tablet is not worn and is thus completely unobtrusive, but the participant has to either hold it with one hand, which leaves only one hand free to perform the task, or put the tablet on a flat surface, which requires head movement to receive instructions. We used a LG G Pad 8.3 in this condition (Tech Specs: 338g, 8.3 niches, 217 × 127 × 8.3 mm).

Paper

In this condition, the instructions are printed in a custom-made paper manual, one page per instruction. To recreate a realistic paper manual form factor, we used the actual manual of the car used in the experiment and we replaced its pages with our printed instruction. To navigate the instructions, participants had to manually turn the pages of the manual. Apart from the navigation, this condition's characteristics are similar to the Tablet condition. The purpose of this condition is to provide a baseline of comparison, since paper manuals are still the most common way of delivering instructions in the maintenance industry.

Tasks and Action Types

We selected eight tasks from the official maintenance inspection checklist used at Firestone, a US auto-care company. Our selection criteria included: 1) Length: The tasks involve a number of steps; 2) Variety: The tasks involve different car components and difficulty levels; and, 3)Representativeness: The tasks can be generalized to other maintenance domains. The descriptions for the tasks were as follows.

Task 1: Engine Oil. Participant checks if the oil level is sufficient using the engine oil dipstick.

Task 2: Fuse (external). Participant pulls out a specific fuse from the exterior fuse box to see if it is blown.

Task 3: Air Filter. Participant checks the condition of the air filter contained inside a housing.

Task 4: Headlight. Participants removes the light bulb from housing of the right headlight and checks if it is burned out.

Task 5: Brake Fluid. Participant checks the brake fluid level.

Task 6: Coolant. Participant checks the coolant level and adds some coolant liquid.

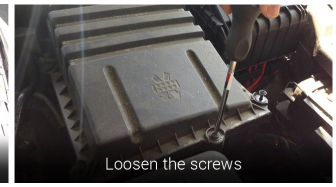
Task 7: Battery. Participant connects a battery tester to the (+) and (-) terminals and reads it.

Task 8: Fuse (internal). Participant pulls out a specific fuse from the interior fuse box to see if it is blown.

Practice Task: Tire Pressure. Participant uses a tire pressure gauge to measure and report the pressure of a target tire.

To design the instructions, we decomposed each task into individual action steps and conveyed each action with one picture and one short sentence written in the imperative. We used actual photographs taken on the car rather than schematics to avoid confusion and we used plain language so novice users could understand. To make sure our instructions were technically accurate, we used the official car manual as a reference and cross-validated with the preventive car maintenance instructions of the CDX Automative Training Textbook (http://www.cdxetextbook.com).

Step2 (Locate Action)



Step3 (Manipulate Action)

Step5 (Compare Action)

Step7 (Read Action)

Figure 3: Instruction example of Task 5 Air Filter.

Previous research [16, 21] categorized maintenance tasks into psychomotor actions, such as manipulation and comparison, and cognitive actions, such as localization and reading. Our task analysis confirmed that Locate-Read-Manipulate-Compare is a recurring pattern of actions in the car maintenance tasks, hence, we classified the task steps into four types of action: Locate, Manipulate, Compare, and Read. Locate involves visual search, typically performed to find a specific car component. Manipulate involves physical manipulation such as unscrewing, lifting and removing. Compare involves visual comparison of what is seen in the real world with what is displayed on the screen, such as assessing the condition of a car component. Figure 3 shows example instructions for Task 5: Air Filter. Step 2 is a an example of a Locate action, Step 3 is a Manipulate action, Step 4 is a Compare action, and Step 7 is a Read action.

We paired these tasks to form four trials (Task1&2 as Trial1, Task 3&4 as Trial2, Task5&6 as Trial 3, and Task 7&8 as Trial4). Our pairing strategy was to include one easy task and one difficult task, (the task difficulty was estimated through pilot testing), and also to avoid pairing two tasks that involve two car components next to each other, so that the participant cannot instantly locate the component for the subsequent task.

Controlled Factors

We counterbalanced the order and learning effects for different technology conditions using a 4 × 4 Latin square (see Table 1). Participants were equally randomly assigned to four testing groups. Every group performed the same sequence of trials, but received a different order of technology condition. This design ensures that each technology condition is tested equally often on a given task.

To control the effects of display size, we designed the instructions to cover the same field of view in all four conditions: 13° horizontally and 7° vertically. Based on this, we calculated the width and height to use at the different eye-to-display distances of the devices. As a result, we used the full screen of Google Glass, 66% of the display surface of the Epson Moverio glasses, and used a 90.8 × 51.1 mm content rectangle on both the tablet and the paper manual as we assume an average reading distance of 40 cm.

Experimental Setup

The study was conducted during the day in summertime at an outdoor parking lot (Princeton, New Jersey, USA). The light level was medium to high depending on the weather and the noise level was low. In addition to the participant, the experiment sessions involved three people: the facilitator, the cameraman, and the wizard. The wizard set up the Wizard of Oz back-end system, prepared the test devices, and initiated the computer responses during the tests when participants gave voice commands.

The apparatus included the car, the four test conditions, and the tools used to complete the tasks. The car was a 2007 Volkswagen Passat. The tools necessary to complete all the tasks were handed to the participant when needed and consisted of a tire pressure gauge, a paper towel, a screwdriver, a bottle of coolant, and a battery tester. Participants were also asked to put black work gloves on both hands before the trials started, and to keep wearing them during the whole trials. The tablet computer was another LG G Pad 8.3 and ran the Wizard of Oz back-end system we programmed. This system was an application to remotely control the test devices through either Bluetooth or Wi-Fi Direct, and to log the navigation events in order to compute automatically the time spent on each task step.

Participants & Procedure

We recruited 12 participants (3 female, 9 male), aged 20 to 27, from Siemens Corporate Technology at Princeton, New Jersey. All participants had an average of 4 years driving experience. They reported to have limited experience with car maintenance. All participants had normal or corrected-to-normal vision except one with red-green color blind. We tested all participants for eye-dominance using the Miles test [23]: seven were right-eye dominant and five were left-eye dominant (the ratio is comparable to the literature[23]).

Each participant was tested individually, and the whole experiment session lasted 45 to 60 minutes. In the first phase, the participant was greeted on the experimental site where the car was parked. A brief was given on the purpose and procedure of the experiment, and an informed consent form was administered. We then collected the demographic information. In the second phase, the participant was randomly assigned to one of the four test groups as shown in Table 1, and received instructions about the system and testing devices. During the practice, the participant was instructed to perform tire pressure check using a pressure gauge (a different tire was used in each test), and report whether the pressure was good or not. During the actual experiment, all participants were told to finish every trial as fast and correctly as possible. After completing all the trials, they were asked to select the most preferred technology, and received a $20 Amazon gift card.

	Trial1 (Task 1&2)	Trial2 (Task 3&4)	Trial3 (Task 5&6)	Trial4 (Task 7&8)
Group1	Paper	Tablet	Eyewear-Peripheral	Eyewear-Central
Group2	Tablet	Eyewear-Peripheral	Eyewear-Central	Paper
Group3	Eyewear-Peripheral	Eyewear-Central	Paper	Tablet
Group4	Tablet	Eyewear-Peripheral	Eyewear-Central	Paper

Table 1. Four testing group received same sequence of tasks, but a different order of technology condition.

Measures

We gathered performance measures, which include completion time and errors. The completion time is the time to complete a step (action), not to complete a whole task. Errors were obtained by comparing the participants' answers regarding the condition of the car components with their actual condition.

RESULTS

Overall, all the participants completed all the tasks quite accurately: ten had no errors, and the other two committed only one error in a single task step. Therefore, these data were not included in the statistical analysis. A 3-way ANOVA (Technology x Task x Action Type) applied to the completion time showed significant main effects for Technology ($F(3, 582) = 3.404$, $p=.017$, $\eta2= .017$, power $(1-\beta)= .768$), Task ($F(7, 582) = 8.862$, $p<.001$, $\eta2= .096$, $(1-\beta)= 1.000$), and Action Type ($F(3, 582) = 38.140$, $p<.001$, $\eta2= .164$, $(1-\beta)= 1.000$). There were also significant two-way interaction effects for Technology x Task ($F(21, 582) = 2.327$, $p=.001$, $\eta2= .077$, power $(1-\beta)= .997$), and Technology x Action Type ($F(9, 582) = 2.808$, $p=.003$, $\eta2= .042$, power $(1-\beta)= .961$), and also the three-way interaction effects Technology x Task x Action Type ($F(63, 582) = 2.469$, $p<.001$, $\eta2= .211$, power $(1-\beta)= 1.000$).

Post-hoc pair-wise comparisons with Bonferroni adjustments on the different Technology conditions showed Eyewear-Central had shorter task completion time compared to Eyewear-Peripheral ($p=.014$), and there were no difference between Eyewear and Non-Eyeware conditions ($p=.344$)(See Figure 4 for details). The main effect of different Task conditions on the completion time was expected because different tasks may vary in terms of difficulty level. Post-hoc pair-wise comparisons with Bonferroni adjustments on the different Task conditions showed that Task 7 (Battery Test) condition had the longest completion time among all the task conditions ($p<.001$). Figure 5 below showed overall the participants spent more time on Task 1 vs. Task 2, Task3 vs. Task4, Task 6 vs. Task 5 and Task 7 vs. Task 8. Although the difficulty order of Task 7 and Task 8 was opposite to our initial assessment, all the other results were consistent with our initial predictions of the difficulty level of each task. This assignment of two tasks with different difficulty levels balanced of the overall task difficulty level in each experiment trial (See Figure 5).

Also as expected, different Action Type conditions yielded the different completion times as shown in Figure 6. Post-hoc pair-wise comparisons with Bonferroni adjustments on different Action Type conditions showed the Locate and Manipulate Action had longest completion time ($p<.020$), and the Read Action had the shortest completion time ($p<.001$). This can be explained by the fact that reading is a relatively straight forward and easy task, requiring less effort for information processing. As there are quite a few machine components and parts under the vehicle hood, it was expected that the participants may need to spend more time searching for the target components or parts.

The interaction effects between different Technology conditions and Action Type revealed an interesting pattern: there were no differences for Read and Compare among different Technology conditions. However, for Locate, the Eyewear-Peripheral condition yielded significant longer completion time than any other conditions (See Figure 7). The interaction effects between different Technology and Task conditions looked complex at first glance. However, further examination of Figure 8 below revealed that for Task 7(Battery Test) and 5 (Brake Fluid), the Eyewear-Peripheral condition produced significantly worse performance in terms of completion time compared with the rest of the Technology conditions, such as Tablet, Paper, or Eyewear-Central. For the remaining six task conditions, there were no differences for the performance among different Technology conditions. One possible explanation is that Task 5 & 7 were the most difficult tasks and required longest completion time as shown in Figure. The design of Eyewear Peripheral was not optimized for the participants to fixate the information for a long time. After reviewing Task 5 & 7, we found that the instructions were ambiguous, whereas the instructions for the other tasks were quite straightforward. For the brake fluid, the participants could not see the level clearly without opening cap of the tank, but this was not explained clearly in the instruction. Eventually, they had to figure this out themselves by opening the cap and looking the fluid level from above. For the battery task, they had to understand that no light on the battery test means the battery is not ok. Again, that was not clear from the instructions.

When we aggregated the data for tasks with clear instructions vs. tasks with ambiguous instructions, we found a significant interaction effect between Eye-wearability and Instruction ambiguity ($F(1, 694) = 3.704$, $p=.055$, $\eta 2= .005$, power ($1-\beta$)= .485) (See Figure 9). It seemed that during these uncertain task situations, the participants in Eyeweable conditions would spend more time on the display information rather than spending time on the machine. It is possible that users developed more reliance on the information perhaps because the display was close to their eyes--a phenomenon that we call the Over-Reliance Effect. This will be further discussed in the next section.

Finally, the three-way interaction effects of Technology, Task, and Action Type are complicated. Most of the effects can be explained by the interaction effects from Technology and Task, and Technology and Action Type. The rest of the effects came from the interactions between Task and Action Type, which is not the focus of study, and therefore will not be further discussed.

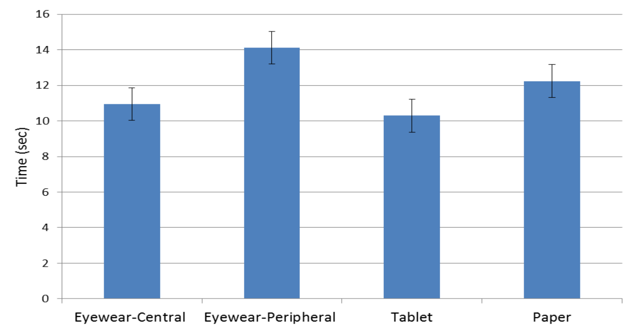


Figure 4. Completion time for different Technology Condition (with S.E. as the error bar).

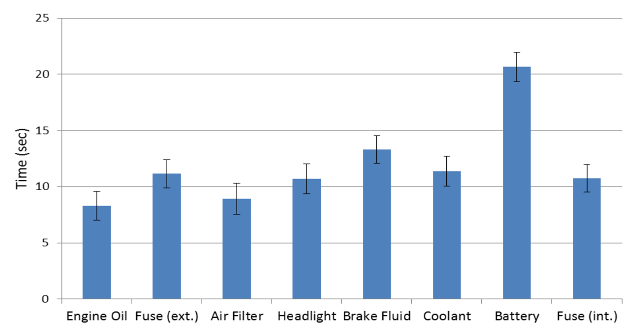


Figure 5. Completion time for different Task Condition (with S.E. as the error bar).

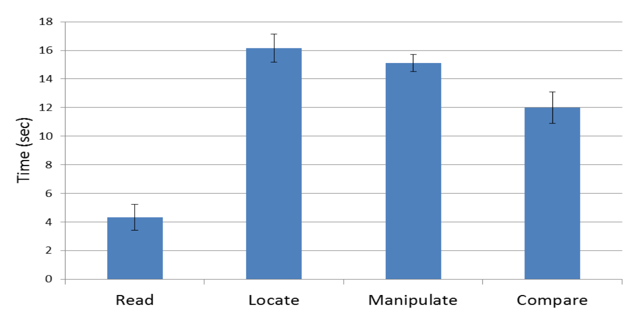


Figure 6. Completion time for different Action Type conditions (with S.E. as the error bar).

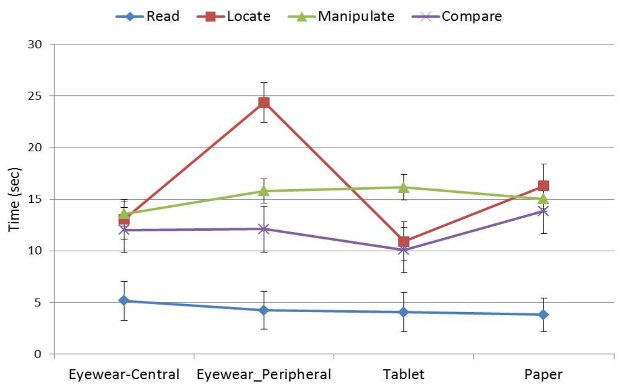


Figure 7. Interaction effects between Technology and Action Type on the completion time (with S.E. as the error bar).

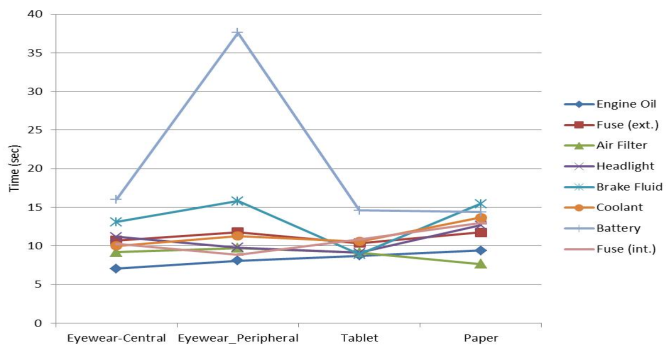


Figure 8. Interaction effects between Technology and Task on the completion time.

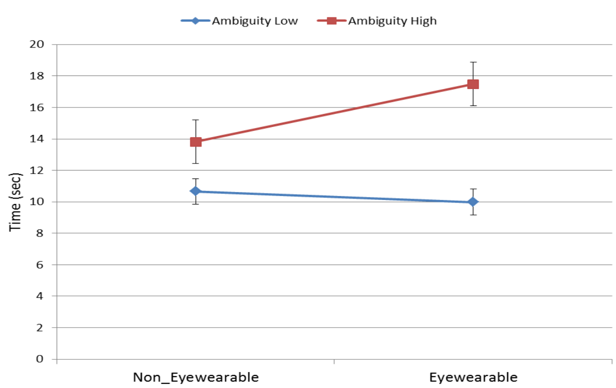


Figure 9. Interaction effects between Wearable and Task Ambiguity on the completion time.

DISCUSSION

Finding 1: Eyewear-Central was faster than Eyewear-Peripheral

Overall, Eyewear-Central yielded a shorter completion time than Eyewear-Peripheral. This can be accounted for by a shorter time to access or process the information in Eyewear-Central compared to Eyewear-Peripheral. Different types of eye movement are involved in accessing information in these two display conditions. Vergence, the typical eye movement involved in the Eyeware-Central, is a frequent and natural eye movement since it is automatically produced when looking at an object at a different distance (e.g. when reading a book at arm's distance and when reading a sign far away). Whereas extorsion, the eye movement in Eyewear-Peripheral, is far less frequent when interacting in the world, and is typically used for quick glances (such as when looking at a rearview mirror while driving) rather than fixations. Initially, we thought that Eyewear-Central would make it harder to focus on the physical world and thus cause some disturbance, however we observed that participants could filter out the information easily, which is consistent with humans' vergence skills [9, 29]. Further analysis showed that the difference in completion time was mostly driven by the Battery and Brake Fluid tasks, and particularly by the Locate actions.

Finding 2: No difference in completion time between Eyewear and Non-Eyewear conditions

To our surprise, we did not detect a significant difference between Eyewear technologies (Eyewear-Peripheral + Eyewear-Central) and Non-Eyewear Technologies (Tablet + Paper). We expected the hands-free characteristic as well as a theoretically shorter information access in Eyewear conditions to have an impact on the completion time. Further analysis of participant behaviors revealed why those factors had a small effect size and were evened out by other factors.

Participants were surprisingly good at adapting the Non-Eyewear technologies to their context or adapting their context to the Non-Eyewear technologies in order to work with their two hands, a phenomenon Mackay refers to as co-adaptation [13]. Participants were sometimes able to perform a step that normally requires the use of two hands with only one hand, for example they could open the coolant bottle by turning the cap with the thumb and index finger while holding the bottle still with the other fingers. When such strategies were not possible, participants quickly found convenient places to put non eye-wearable devices, such as on the car frame or on top of the engine. In those cases, the instructions are only a head movement or upper-body movement away, versus an eye movement or eye convergence away in Eyewear conditions. This limits the difference in access time between the two to less than 1 second. Although not statistically significant, we do see a trend where Eyewear may be faster than Non-Eyewear on the Fuse (int.) Task, which required more effort to read the instructions on non eye-wearable devices because participants had to crouch down or to bend sideways from the driver's seat. Other machine maintenance environments may be more restrictive than ours; hence future work will investigate the interaction effects with diverse machine installations and body position requirements (e.g. lying).

In addition, the semi-transparent screen on eye-wearable technologies also made it harder to see the texts and pictures clearly in both Eyewear conditions, to the extent that sometimes participants put one of their hands behind the screen to provide a dark background since they were wearing black gloves. The hands-availability factor was therefore evened out between Eyewear and Non-Eyewear when this happened. Eye-wearable technology makers may consider using different opacity levels, especially for the Eyewear-Peripheral kind of devices since users can still see the physical world right in front of them regardless of the opacity of the screen.

Finding 3: Eyewear-Peripheral was slower than any other technology on Locate actions and on certain tasks.

As mentioned earlier, analysis of the three-way interaction effects revealed that the longer time of Eyewear-Peripheral on the Battery and Brake Fluid tasks was mainly spent on Locate actions. As it happens, the instructions corresponding to Locate actions on those tasks were harder to interpret than on the other tasks. This can be explained by the Multiple Resource Theory [34]. For most actions involved in the trials, participants could multi-task between performing the maintenance tasks and processing the instruction displayed by the technology, but this was not possible with Locate actions: processing the instruction and locating the objects of interests both require the visual attention, thus causing any time spent processing the instruction to accumulate in the total completion time.

But why was this visual processing much longer when they used the Eyewear-Peripheral technology? Two unique characteristics of the Eyewear-Peripheral display cause users to be much less efficient at processing the information displayed for a sustained period of time compared to the other technologies: the first characteristic is the position of the display relative to the user's eye, and the second characteristic is that it is monocular.

First, the Eyewear-Peripheral display is eccentric to the normal line-of-sight (upper right hand corner), and therefore suboptimal for eye fixation. When humans need to look at an object in the physical world for an extended period of time, in order to align the object with their line-on-sight, it is much more natural for them to move their head, to move their body, or to move the object with their hands, rather than to move their eyes at a high angle. The same holds true when humans need to look at a digital display. Because human visual acuity is best in the fovea (in the center of the retina), users cannot rely on their peripheral vision to see the information well [19]. With Eyewear-Peripheral, they have to move their eyes to align the fovea with the display. This puts the eye at an unnatural angle and requires a significant muscular effort to hold that position. Human vision is less efficient at fixating eccentric targets compared to eye-centered targets, possibly because of microsaccades, which are involuntary and necessary eye movements during fixation, but cause the eye to diverge more from the target the more it is eccentric. The decreased visual acuity in people with strabismus, as shown by [19], is evidence that our eyes perform better when looking straight. As a result, the longer participants had to fixate the display to process the information, the more their performance with the Eyewear-Peripheral display was decreased.

Second, the Eyewear-Peripheral display is monocular, and therefore conveys visual information less accurately than binocular displays. This is mainly due to three phenomena of human binocular vision: binocular summation, binocular interaction, and binocular rivalry. Binocular summation enhances visual perception by combining the information received from the left and the right eyes. It notably increases visual acuity (by 33%) and spatial contrast sensitivity (by 40%) [29]. Binocular interaction means that the state of one eye, such as its focus and pupillary diameter, influences the state of the other so that they even out. Binocular rivalry is a phenomenon that occurs when two different images are presented to each eye simultaneously: perception alternates between the two images. Due to this interference, it is harder to focus on the image presented to one eye when the other eye perceives a different image, unless closing the other. This can explain why participants were sometimes closing the left eye when looking at the Eyewear-Peripheral display. Loss of binocular summation and interference due to binocular interaction and binocular rivalry all contribute to degrade visual perception and further reduce the ease of processing Locate instructions in detail with Eyewear-Peripheral.

Finding 4: Eyewear conditions yielded longer completion times on ambiguous instructions

During the trials, we observed that participants got stuck on three steps, one in the Brake Fluid task, two in the Battery task. Participants were puzzled because the instructions of those steps, did not provide sufficient information to complete the steps. Dealing with this ambiguity required an understanding that they could only draw from the physical world and from their own knowledge. This was reflected in the completion times of those steps: they were significantly and consistently longer than the other completion times, but to a greater magnitude with the eye-wearable technologies compared to the non eye-wearable technologies.

Eye-wearable technologies seemed to decrease participants' ability to think by themselves: we call this the over-reliance effect. We interpret it as the result of an unprecedented physical proximity of the technology to the human: because the technology is so close and is always present, it has a greater power and influence on the user, increasing the extent to which the user trusts the technology. Our results are not sufficient to conclude that such an effect indeed exists, however, similar effects have already been observed in previous work. In wearable computing, Ockerman & Pritchett[17] found that avionics performing a preflight inspection detected less faults and neglected to inspect items more often with a head-mounted display vs. doing the inspection by memory when these faults were not mentioned on the head-mounted display. Over-reliance on automation is also a very well-known and well documented issues [11, 18], for example on driver assistance systems. In social psychology, the construal level theory [32] related psychological distance with people's capability of abstracting: the closer the object is, the more concretely it is thought of. In the similar way, Williams [35] conducted several studies showing that physical distance influences people's thoughts and judgments: when reading a book or an article at a closer distance, people were less detached. These findings support our theory that the close proximity of eye-wearable technology affects user's level of reliance on the technology. This over-reliance effect requires more investigation as it could have dramatic implications.

Finding 5: Eyewear-Peripheral was preferred to all the other technologies.

User preference ratings showed that Eyewear-Peripheral was most preferred (6 participants most preferred Eyewear-Peripheral, 3 for Eyewear-Central, 2 for Tablet, and 1 for Paper). Their justifications included "it is hands-free", "it is light and comfortable", "it is unobstructive and non-distracting", and "it is convenient". The biggest complaint was that they "didn't like to look in the corner" and "didn't like to change their vision from up to down because it is awkward and hard to adapt". Comments for Eyewear-central include "heavy", "uncomfortable", "don't like having the information in front of me always". Comments for both Tablet and Paper included "see things clearly", "no need to adapt vision", "easy to carry around", "worried about dropping it now and then", "could get dirty", "everyone knows how to use", but "annoying to turn the pages(for Paper).

CONCLUSION

This paper presented a method to disaggregate the effects of technological factors, tasks, and action types on user performance of procedure following in machine maintenance. We devised an experimental setup that is easily replicable and ecologically valid, and an experimental design that isolated the effects of essential eye-wearable attributes by controlling the effects of interaction modality, display size, and instruction design, and that maximized the amount of measures gathered per participants. Our study results showed: In terms of step completion time, the peripheral eye-wearable display was significantly longer than the central eye-wearable display overall and was significantly longer than any other technology in Locate actions, mainly because its position relative to the eye is unsuited for long fixations and also because it is monocular; The hands-free advantage of eye-wearable technology was not observed because participants were skilled at adapting the non eye-wearable technology and the situation to work with their two hands; The eye-wearable technologies yielded significantly longer completion times when instructions were ambiguous, suggesting that the physical proximity of eye-wearable technology to the user may influence the trust level, causing the over-reliance effect.

Additional systematic empirical work is needed in order to understand how eye-wearable technology can be leveraged to benefit workers significantly in industrial settings. We hope the method introduced in our study will inspire researchers to conduct other studies and generate new scientific knowledge. Future work should investigate factors that our study didn't cover, such as different display sizes and positions, different interaction modalities and interaction techniques, and different instruction workflow designs. Different task environments, different level of expertise with the tasks, and different performance and experience measures should be also used. To study the effects of tasks in a more meaningful way, a comprehensive taxonomy of machine maintenance tasks should also be provided. In addition, we would like to raise attention on some issues we discovered that were largely overlooked by previous work and need deeper examination, namely: the effects of the eye-wearable display position relative to the user's eye(s), the effects of monocular vs. binocular display, how these effects interact with different visual perception phenomena, and the potential over-reliance on eye-wearable technology.

ACKNOWLEDGMENTS

We thank all the reviewers for their helpful comments and suggestions. We thank Yuan (Emma) Meng for creating images, Kai-Sheng (Roy) Chuang and Ying Zhu's support for the study, and Bill Sherman for proofreading the paper.

REFERENCES

1. Aleksy, M., Rissanen, M. J., Maczey, S., & Dix. M. Wearable computing in industrial service applications. *Procedia Computer Science*, 5:394-400, 2011.
2. Billinghurst, M., Bowskill, J., Jessop, M., & Morphett, J. A wearable spatial conferencing space. In *Proc. 2nd International Symposium on Wearable Computers*, 76-83. IEEE, 1998.
3. Caudell, T.P., & Mizell, D.W. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proc. 25th Hawaii International Conf. on System Sciences*, 659-669, 1992.
4. Dey, A. K., Salber, D., Abowd, G. D., & Futakawa, M. The conference assistant: Combining context-awareness with wearable computing. *In Proc. 3rd International Symposium on Wearable Computers,* 21-28. IEEE, 1999.
5. Friedrich, W. ARVIKA-augmented reality for development, production and service, *In Proc. ISMAR*, 3-4, 2002.
6. Gartner Says Smartglasses Will Bring Innovation to Workplace Efficiency. http://gtnr.it/1qqbXce.
7. Goose, S., Sudarsky, S., Zhang, X..& Navab, N. Speech-enabled augmented reality supporting mobile industrial maintenance, *Pervasive Computing*, IEEE, 2(1): 65-70, 2003.

8. Henderson, S. & Feiner, S. Augmented Reality in the Psychomotor Phase of a Procedural Task, In *Proc. of IEEE International Symposium on Mixed and Augmented Reality*, 191-200, (2011).
9. Hung, G. K. ,& Ciuffreda, K. J. *Models of the Visual System*. Springer Science, 2002.
10. Kraut, R. E., Fussell, S. R., & Siegel, J. Visual information as a conversational resource in collaborative physical tasks. HCI, 18, 13-49, 2003.
11. Lee, J. D., & See, K. A. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50-80, 2004.
12. Machinery Maintenance & Heavy Equipment Repair Services Market Research Report, 2014. http://bit.ly/1kVmv6N.
13. Mackay, W. E.. Users and customizable software: A co-adaptive phenomenon. PhD thesis, 1990.
14. Mann, S. Wearable computing. *The Encyclopedia of HCI*, 2nd Ed., 2013.
15. Mann, S., Lo, R. C., Ovtcharov, K., Gu, S., Dai, D., Ngan, C., & Ai, T. Realtime high dynamic range video for eyetap wearable computers, fpga- based seeing aids, and glasseye. In *Proc. 25th CCECE*, 1-6, 2012.
16. Neumann, U., & Majoros, A. E. Cognitive, performance, and systems issues for augmented reality applications in manufacturing and maintenance. In *Proc. VRAIS*, p4-12, 1998.
17. Ockerman, J. J., & Pritchett, A. R. Preliminary investigation of wearable computers for task guidance in aircraft inspection. In *Proc. 2nd International Symposium on Wearable Computers* 33-40. IEEE, 1998.
18. Parasuraman, R., & Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors,* 39(2):230-253, 1997.
19. Pineles, S. L., Velez, F. G., Isenberg, S. J., Fenoglio, Z., Birch, E., Nusinowitz, S., & Demer, J. L. Functional burden of strabismus: Decreased binocular summation and binocular inhibition. *JAMA ophthalmology*, 131(11):1413-1419, 2013.
20. Reiss, M., & Reiss, G. Ocular dominance: some family data. *Laterality: Asymmetries of Body, Brain and Cognition*, 2(1):7-16, 1997.
21. Richardson, M., Jones, G. & Torrance, M.. Identifying the task variables that influence perceived object assembly complexity, *Ergonomics*, 47(9), 945-64, 2004.
22. Rhodes, B. J. The wearable remembrance agent: A system for augmented memory. *Personal Technologies*, 1(4):218-224, 1997.
23. Siegel, J., & Bauer, M. A eld usability evaluation of a wearable system. In *Proc. 1st International Symposium on Wearable Computers* 18-22. IEEE, 1997.
24. Siewiorek, D., Smailagic, A., Bass, L., Siegel, J., Martin, R., & Bennington, B. Adtranz: a mobile computing system for maintenance and collaboration. In *Proc. 16th International Symposium on Wearable Computers,* 25-25, IEEE, 1998.
25. Smailagic, A., & Siewiorek, D. Application design for wearable and context-aware computers. *Pervasive Computing*, IEEE, 1(4):20-29, 2002.
26. Smart Glass Market by Technology -- 2020. http://bit.ly/1thCiOK.
27. Starner, T. The challenges of wearable computing: Part 1. *IEEE Micro*, 21(4):44-52, 2001.
28. Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R. W., & Pentland, A. Augmented reality through wearable computing. *Presence: Teleoperators and Virtual Environments*, 6(4):386-398, 1997.
29. Stidwill, D., & Fletcher, R. *Normal Binocular Vision: Theory, Investigation and Practical Aspects*. John Wiley & Sons, 2010.
30. Thompson, C., Najjar, L. J., & Ockerman, J. J. Wearable computer based training and performance support systems. In *Proc. the Interservice/Industry Training, Simulation & Education Conf.*, 1997.
31. Träskbäack, M. & Haller, M. Mixed reality training application for an oil refinery: user requirements. In *Proc. of VRCAI*, 324-327, 2004.
32. Trope, Y., & Liberman, N. Construal-level theory of psychological distance. *Psychological review*, 117(2):440, 2010.
33. Weaver, K. A., Baumann, H., Starner, T., Iben, H., & Lawo, M. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proc. of the SIGCHI*, 1695-1704, 2010.
34. Wickens, C. D. Processing resources in attention. (Parasuraman R, Davies R), editors. *Varieties of attention*, 1984.
35. Williams, L. E. & Bargh, J. A. Keeping one's distance the influence of spatial distance cues on aect and evaluation. *Psychological Science*, 19(3):302-308, 200.
36. World Vehicle Population Tops 1 Billion Units WardsAuto. http://bit.ly/1jugNFm.