# What's Next? Designing Head-Mounted Displays for Procedural Task Support

**Xiang Zhi Tan, Steven Johnson, and Bilge Mutlu**
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706 USA
xtan, sjj, bilge@cs.wisc.edu

## ABSTRACT

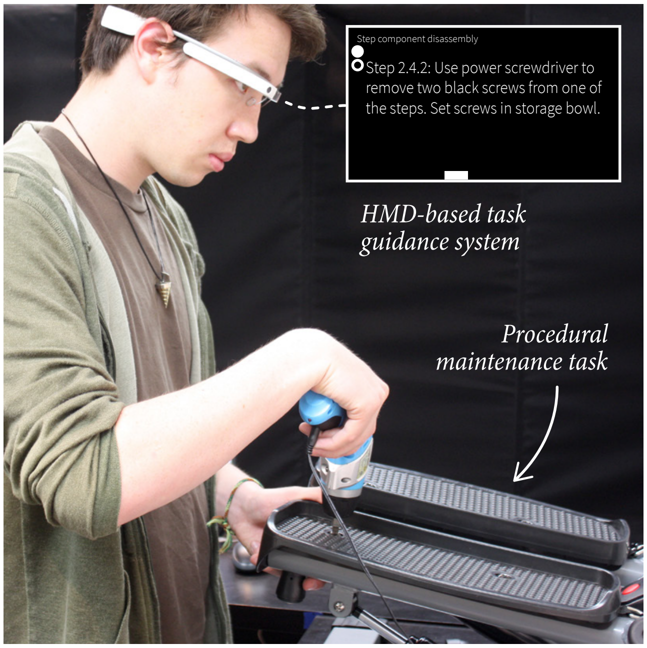
Many day-to-day activities, from cooking to repairing an appliance, involve following specific *procedures*. Performing unknown or infrequent procedures, such as cooking a recipe for the first time, requires following instructions provided by an expert. The vast majority of such instructions are in the form of textual lists presented on paper or on a screen. While this presentation offers ubiquitous use, it introduces attentional and physical demands due to frequent context switching. In this paper, we explore how lightweight head-mounted displays may offer such support and what interaction styles may provide the best support. We designed four variations of such a support system involving combinations of *display-only and display-and-voice output* and *input through voice and touch-gestures*. We evaluated these designs against a computer-based checklist in a maintenance task. While the HMD designs provided only marginal performance gains, participants found them to be more suitable for context switching.

### Author Keywords

Procedural task support; head-mounted displays; voice input and output

### ACM Classification Keywords

H.5.1. **Multimedia Information Systems**: *Artificial, augmented, and virtual realities, Audio input/output*; H.5.2. **User Interfaces**: *Graphical user interfaces (GUI), Input devices and strategies (e.g., mouse, touchscreen), Screen design (e.g., text, graphics, color), Voice I/O*



Step component disassembly

Step 2.4.2: Use power screwdriver to remove two black screws from one of the steps. Set screws in storage bowl.

**Figure 1. In this paper, we explore how lightweight head-mounted displays may provide guidance in procedural tasks. We designed and developed an automated task guidance system using the Google Glass platform and Procedure Representation Language (PRL) and evaluated different user input/output configurations in a maintenance task.**

## INTRODUCTION

Day-to-day activities from cooking a meal to repairing a piece of machinery, completing chemistry labs to performing a surgical operation, and assembling a piece of furniture to navigating through a city involve following specific "procedures" that ensure the successful and efficient completion of the task. When procedures are complex, unfamiliar, or infrequently performed, individuals rely on a set of instructions provided by an expert. These instructions can be presented in person, such as a physical therapist guiding a patient through exercises, through pre-recorded and authored video, such as "how-to" videos available online, or in text, such as recipes provided in a cookbook. While all of these forms of delivery are common, textual descriptions are ubiquitous due to their flexibility and accessibility.

Despite the ubiquity of their use, textual instructions are also the most difficult to use while performing a physical task due to physical and attentional demands placed on the performer. These demands result from the need to repeatedly switch contexts between the task and the instruction, the use of hands to navigate within the instructions, such as holding a piece of paper, turning a page, or scrolling down in a list, and the often challenging environments in which procedures are performed, such as under the hood of a car, a surgical opening, or a field setting with high wind. How could technology relieve performers of these demands and facilitate easy, effective, and more accurate following of task procedures?

Prior work has explored how mobile and wearable computers can offer task guidance in specific settings, such as performing aircraft inspection [13] and maintenance [17], addressing emergencies in nuclear power plants [15], picking orders at a warehouse [19], completing lab procedures in a wet lab [16], and performing car maintenance [20]. These studies found that the systems developed often interfered with the performance of the procedure [11], lacked information on task structure and progress [8, 9], and added additional navigation load [13]. Therefore, the potential benefits of the use of mobile and wearable computers for procedural tasks and what specific design elements may best support performers remain unknown.

In this paper, we investigate whether or not head-mounted displays can relieve performers of the demands placed by the use of textual instructions and explore the design space for how such systems may offer procedural task guidance. We designed a core task-guidance system built on Procedural Representation Language (PRL) [7], an XML-based representation designed for task instructions, and a hierarchical visual representation to provide the user with an awareness of the structure of and their progress in the procedure. Our exploration also involved creating visual and voice-based presentation of instructions and gesture- and voice-based user input for navigation. We evaluated four versions of our design that combined different forms of presentation and input, comparing them against a baseline checklist (also based on a PRL representation) presented on a computer.

In the remainder of this paper, we review related work on task-guidance systems, describe the design and implementation of our system, provide detail on the user study, present our findings from the study, and discuss these findings, the limitations of the research, and our future work.

## RELATED WORK

A review by Ockerman and Pritchett [11] refers to computer systems that provide workers with guidance and serve as a reference for task procedures without any sensing of the environment "task guidance systems." Previous research on task guidance system and their use fall under three main categories: (1) designs of task guidance systems to support procedures in specific domains, (2) comparisons of computer-based guidance systems against conventional procedure guidance, and (3) studies that explore the most effective ways to present procedure information. The paragraph below briefly outline key studies that fall under these categories.

### Designs of Task Guidance Systems

A large number of task guidance systems were developed in the late 1990s using mobile and wearable computing platforms due to the primarily field settings for which they targeted. For instance, Roth [15] developed a system to support emergency procedures followed at nuclear power plants, focusing particularly on aiding in higher-level decision-making and improving situation awareness, response planning, and error recovery. A wearable computer developed by Ockerman and Pritchett [13] consisted of a display, microphone, and earphone and guided users through preflight inspection protocols. The system supported hands-free operation and offered a checklist to keep track of completed procedure steps. Siegel and Bauer [17] designed a wearable system consisting of a touchpad and stylus to support procedures in avionics maintenance tasks. More recent work by Scholl et al. [16] involved the development of a task-guidance and -recording system using the Google Glass platform and an additional wrist sensor to be used in a wet lab. The system provided users with simple instructions and enabled them to navigate between instruction steps, record their work, and "check" the step when they were done.

In addition to computer-based systems, researchers have also investigated the use of augmented reality for task guidance. For instance, several medical applications involved augmented-reality systems that helped doctors with treatments and surgeries [4, 5, 14]. While these systems do not provide users with direct instruction, they are created to provide medical personnel with auxiliary assistance by overlaying the information in the user's field of view. Augmented-reality-based task-guidance systems have also been explored in maintenance [6] and assembly domains [1]. An example of such systems is a light-based task-guidance system developed by Sodhi et al. [18] that guided the user's hand motions by projecting goals onto the user's hand.

In their review, Ockerman and Pritchett [11] identified a number of challenges that were common to the design of these systems. First, they argued that systems that are designed poorly can interfere with performing the procedure, resulting in workers abandoning the use of the system. The prototype built by Ockerman and Pritchett [13] illustrated this problem; the system was too clunky for practical use, requiring workers to carry heavy battery packs on their waists. Second, poor design can also cause over-reliance on the system when procedures are erroneous or inappropriate. Comparisons of task guidance systems against paper-based protocol guidance have found that performers rely more on the procedure presented by the computer than paper-based procedures, highlighting the potential dangers of blindly following incorrect procedures [13]. Finally, interruptions while performing procedures may result in the worker skip important steps or get lost. Studies in the aviation [8] and spaceflight [9] settings found interruptions during flight procedures to be extremely disruptive to ongoing procedure performance.

### Benefits of Task Guidance Systems

Prior research on task guidance systems include several studies that sought to establish the effectiveness of these systems in supporting procedural tasks by comparing them to conventional procedure support, such as the use of paper-based instructions. These studies concluded that task guidance systems engaged the workers in the procedure more, although failing to improve their performance and often slowing them down. Siegel and Bauer [17] found that the users of the task guidance system were more thorough in completing the procedure steps than users of paper-based procedures. Compared to paper, these systems also lacked the ability to provide workers with an awareness of the overall structure of the procedure and their progress in it. Ockerman et al. [12] compared the use of paper- and computer-based instructions in an origami task and found workers to repeatedly review upcoming steps in the procedure, for which paper-based instructions provided better support, as instructions were printed on a single page. They suggested that task guidance systems must be designed to support the need to "look ahead" in the procedure. Siegel and Bauer [17] similarly reported users expressing the need for being able to see an overview of the procedure being performed.

A recent study by Zheng et al. [20] compared instructions provided by different forms of HMDs, particularly HMDs that provided "peripheral" and "central" views, against tablet-computer- and paper-based instructions in a car-maintenance task. They found no task-performance differences between HMDs and tablet- or paper-based instructions and that the use of the center-view HMD resulted in better performance than the peripheral-view HMD. The authors attribute the lack of differences to participants' ability to place the tablet or the paper near the task space, which provided participants with hands-free access to instructions, and their adaptiveness to task demands, such as completing task steps that require both hands with one hand.

### Information Presentation in Task Guidance

Another line of work on computer-based task guidance studied how these systems must present information to most effectively support procedure performance. Michas and Berry [10] conducted investigated the effectiveness of different forms of information presentation for learning a first-aid task, comparing text, line drawings, the two combined, video, and video stills. They found that the combination of text and line drawings and video resulted in significantly faster and more accurate learning of the task. More recent work by Weaver et al. [19] explored how head-mounted displays may assist warehouse workers in picking orders, replacing textual, list-based pick orders with visual schematics shown on the HMD display and found HMD-based information display to result in the faster task times compared to visual schematics shown on paper, list-based orders shown on paper, or audio instructions. While both studies highlight the importance of choosing appropriate information representation for task guidance systems, the applicability of their findings to longer and more complex task procedures is unknown.

Prior work offers several design guidelines based on "lessons learned" from the development, deployment, and testing of the prototypes resulting from studies described above as well as on findings from experimental studies. These guidelines include the need to reflect task structure in the system in order to improve worker awareness of the procedure and offer mechanisms to navigate this structure in order to review upcoming task steps, maintain an awareness of task progress, and mitigate the impact of interruptions. Additional guidelines include the need to provide workers with the ability to drill down for information as needed during a task step, while maintaining an awareness of the overall task structure, and with rationale on the goals of the procedure to get worker buy-in and mitigate any potential noncompliance with procedure instructions [11].

The guidelines outlined above serve as critical specifications that our designs must meet in order to provide hands-free support for longer and more complex task procedures. In the next section, we describe our system design as informed by these guidelines.

## SYSTEM DESIGN

In the following section, we discuss the interaction design and system implementation of our system. At a high level, the system consists primarily of a head-mounted display (HMD) that loads a pre-specified Procedure Representation Language (PRL) document, parses textual instructions and associated multimedia materials that are retrieved in real time from a connected multimedia server, and makes these materials availabile to the user.

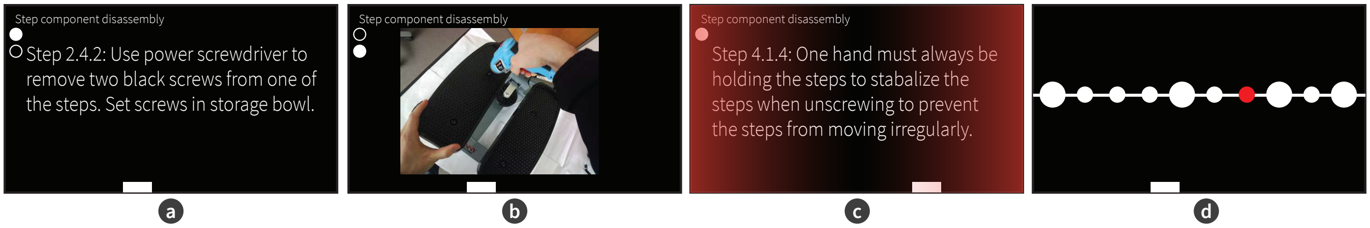
### Procedure Representation Language (PRL)

PRL is a XML-based execution language developed and primarily used by NASA in spaceflight missions to represent procedures that contain operation instructions [7]. PRL provides the operator with a list of instructions that are encapsulated in steps each of which specifies a unique goal to accomplish. PRL allows different types of instructions based on the task. A commonly used instruction in our task would be a "manual" instruction where the operator would need to carry out a manual physical task with a part or a tool, such as unscrewing a screw or pushing a button. PRL also allows us to specify prompts and action requests, such as warnings, cautions, and prompts to take a picture. A PRL code snippet used in the study is shown in Figure 2.

### HMD Platform

```
...
<prl:Step stepIdentifier="step1">
   <prl:StepTitle>
     <prl:StepNumber>1</prl:StepNumber>
     <prl:Text>Obtain exercise equipment and
     tools</prl:Text>
   </prl:StepTitle>
   <prl:ManualInstruction executionMode="human"
   instructionIdentifier="instr03">
     <prl:Description>
       <prl:Text>Ensure stepper exercise
       equipment is at the workstation.
       </prl:Text>
     </prl:Description>
     <prl:Number>1.1</prl:Number>
     <prl:InstructionMessage>
       <prl:Text>Manual</prl:Text>
     </prl:InstructionMessage>
   </prl:ManualInstruction>
   <prl:Image infoIdentifier="info01">
     <prl:ImageReference
     source="../20150220_132622_118.jpg"/>
     <prl:ImageTitle>
       <prl:ImageNumber>1</prl:ImageNumber>
       <prl:Text>Image</prl:Text>
     </prl:ImageTitle>
   </prl:Image>
   ...
```

**Figure 2. An example PRL code snippet used in the study.**



Step component disassembly

Step 2.4.2: Use power screwdriver to remove two black screws from one of the steps. Set screws in storage bowl.

Step component disassembly

Step component disassembly

Step 4.1.4: One hand must always be holding the steps to stabalize the steps when unscrewing to prevent the steps from moving irregularly.

**Figure 3. (a) The design of a standard instruction card. The bubbles on the left serve as a discrete vertical scrollbar indicating what level the user is currently at, and the bottom continuous scrollbar shows overall task progress. (b) Additional context on the task step, such as a photo description of the step, displayed at a lower level. (c) An example of warning card. (d) Example of the high-level context view. Leftmost circle indicates the task step that the user is currently performing.**

Google Glass[1] was selected as the HMD used in the study due to its availability and its potential to support hands-free task guidance in field settings. The Google Glass platform was develop by Google, Inc. in 2013 and since became an icon for modern HMD technologies. The platform uses Glass OS (XE) as its operating system, an extension of the Android Operating System version 4.4 used on smartphones. Our task-guidance application was develop using the Glass Development Kit (GDK), which is included in the Android version 4.4 SDK.

## Application

Our design follows the design patterns developed for Google Glass where content is presented on individual screens called "cards." Our system parses the PRL document and generates an individual card for each instruction. Figure 3 illustrates different card designs. Each card displays the step title in small text on the upper left corner of the screen in order to give the user an idea of the goal the instruction is intended to achieve. The middle of the screen is reserved to display the main card content. If the card contains an instruction, the step number and the instruction text are loaded in this area. If the instruction contains more text than what would fit on the screen, an additional card with the overflow text is produced. If the card is also a warning or a caution, the edges of the screen are highlighted with red color for warnings and yellow color for cautions. Photos included in instruction steps are interpreted by the system as additional content intended to provide clarification to the preceding instruction, such as showing the user which screw should be unscrewed or where the part to be manipulated is.

Our system also supports taking photos with using the HMD's integrated camera when the parsed instruction includes a "camera" tag. The system displays a viewfinder in the middle of the screen to help the user position the HMD appropriately and allows the capture of the displayed scene using a single *tap* gesture. After the first capture, the user can retake the photo by tapping again.

Each instruction card is placed on a hypothetical 2D plane where the next instruction is on the right of the current card and the previous instruction is on the left. The improve user awareness of their position in this 2D plane, we placed a scrollbar at the bottom of the screen to show where the user currently is in the procedure. Additional parsed content for each card is placed below the current card in the plane and distributed across several "levels" of additional content. To indicate to the user that there is additional content, we designed a bubble-shaped discrete scrollbar system and placed it on the upper left corner of the screen where the bubble for the current level is filled up. Figure 4 illustrates our navigation system.

Our task guidance system includes an additional screen called "context view" designed to display what types of instructions are involved in upcoming steps. An example portion of the context view is shown in the Figure 3 (d). Instructions are represented as small circles, and new steps are presented as large circles. The color of the circle changes based on the type of information; green indicates *action*, such as taking a photo, and red and yellow represents warning and caution, respectively. Our goal with the context view is to provide users with additional information about upcoming steps, since findings in prior work indicates a strong need for supporting "looking ahead" [12].

## Interaction

A key goal of our study is to gain a better understanding of what types of HMD interaction styles best support task guidance. Therefore, we designed our system to allow two

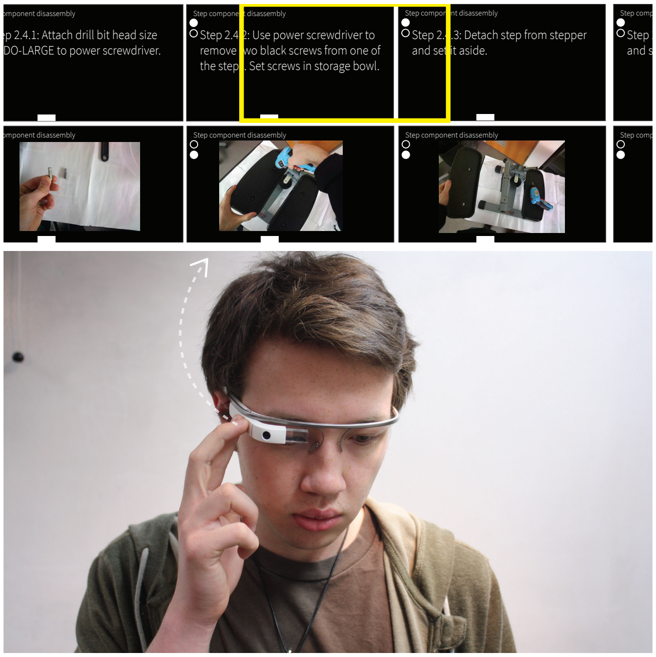


**Figure 4. As users navigate the interface, the 2D array of cards moves until the card that the user wishes to see aligns with the viewing window (shown as a yellow box). The two rows of cards indicate different "levels" of information on the task step.**

---

[1] https://developers.google.com/glass/

alternative methods of user input. The first input method is through touch gestures on the touchpad located on right side of the HMD. The application accepts five commands: *swipe up*, *swipe down*, *swipe right swipe left* and *single tap*. The first four commands are used in navigating the 2D plane of instruction content, and the single tap command is dedicated to enable the user to take a photo. The second input method involves the use of voice commands. Each voice command is directly mapped to a gesture command. For instance, to execute the *swipe up* command, the system expects the user to say "Up, Glass" instead. The addition of the word "Glass" at the end of each word is to prevent the system from recognizing command words when they are used in other forms of speech, such as an ongoing conversation with a collaborator. We implemented these two alternative input methods as separate programs, so that we could initialize the desired version for experimental testing, although we envision that the future version of our system will simultaneously support both methods.

In addition to the two input methods, we built the capability to read instructions displayed on the HMD screen through the integrated bone-conduction speakers of the Google Glass platform. If the text overflows into the next card, the system automatically scrolls to the next card and repeats this process until the end of the instruction.

## HYPOTHESES

To formalize our predictions of the improvements that HMD-based task-guidance system will offer over conventional task-guidance methods, we developed a set of hypotheses. Each hypothesis includes two parts for predictions regarding task performance and user-experience outcomes.

> *Hypothesis 1 (a)* – Participants using the HMD will perform the task faster than those using the computer system.

We believe that the distance between the task space and the computer will result in delays, similar to the "homing" effect included in the Keystroke-Level Model (KLM) [3], and this context switching will result in increased task times with the computer.

> *Hypothesis 1 (b)* – Participants will evaluate the HMD more favorably than the computer system.

We believe that the HMD will result in an overall more positive user experience, as it will reduce attentional and physical demands on the user.

> *Hypothesis 2 (a)* – Users of the HMD with voice control will perform the task faster than those using gesture-based input.

Because using gestures on the touchpad would require moving one hand away from the task space, we expect voice-based input to benefit from both hands remaining in the task space. Furthermore, voice control will also allow users to look ahead for upcoming instructions while their hands are occupied with performing the current instruction.

> *Hypothesis 2 (b)* – HMD users will evaluate voice input more favorably than gesture-based input.

We expect the reduction in physical demands described above to result in a more positive experience.

> *Hypothesis 3 (a)* – Users of the HMD with read aloud will perform the task faster than those without the read aloud..

We expect the read-aloud feature to enable participants to focus on the current task without moving their gaze from the workspace to the HMD. This feature will allow the user to listen to the next instruction while working on the current step, serving as a different form of look ahead.

> *Hypothesis 3 (b)* – HMD users will rate the system with the read-aloud feature more favorably than the system without the feature.

We expect the ability to hear upcoming instructions and maintain visual attention on the task space to result in an improved user experience.

## STUDY DESIGN

To test our hypotheses, we designed a-two×two-plus-baseline (a total of five conditions) between-subjects-design study where the input method (voice vs. gesture) and the presence of the read-aloud feature (absent vs. present) served as the two independent variables. The baseline condition involved the use of a computer with an interface that displayed a checklist generated from the PRL document used by the HMD. The checklist, generated by a program called *Pride View*, displayed instructions defined for the procedure and the ability to "check" off completed instructions.

### Participants

We recruited a total of 45 participants for the study. However, due to the technical difficulties, such as video corruption and system failure, data from eight participants were excluded, resulting in data from a total of 37 participants (18 females, 19 males) with an age range of 18–32 ($M = 21.78, SD = 3.75$) data in the final analysis. The breakdown of participant for each condition was as follows: baseline computer condition (3 females, 4 males), Gesture and no read-aloud (4 females, 4 males), gesture and read-aloud (3 females, 3 males), voice and no read-aloud (4 females, 4 males) and voice and read-aloud (4 females, 4 males).

### Task

To evaluate our hypotheses, we created a task procedure that involved disassembling and reassembling a piece of exercise equipment—a small stair stepper. The process involved removing 10 parts and 10 screws from the stepper and placing them on the table next to stair stepper. The participant was also given a power screwdriver and wrench to assist them in the task. First, the participant was instructed by the system to dissemble the stair stepper through 24 instructions given by the system. After dissembling the equipment, the participant was then instructed to either use the HMD system to take four different pictures of the dissembled equipment or using a mobile phone to capture the photo if they were in the computer condition. After taking the pictures, the system provided another 45 instructions on how to reassemble the stair stepper.

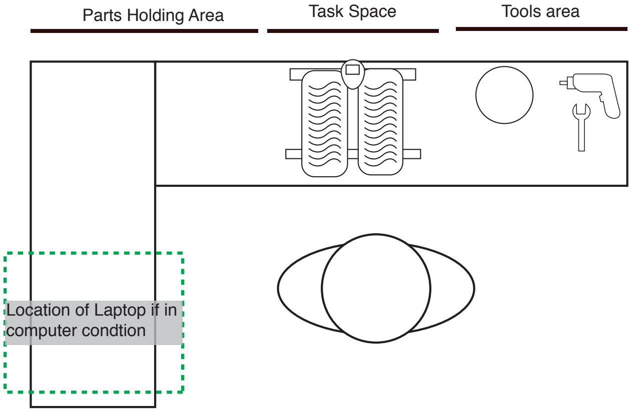


**Figure 5. An illustration of the physical setup of the workspace.**

## PROCEDURE

After obtaining informed consent from the participant, the experimenter introduced the system the participant would be using in the study. If the participant was in the computer condition, the participant was shown the Pride View interface on an 11-inch laptop and taught how to navigate the interface with the mouse provided. They were also shown how to operate the smart phone that was used in the study as a camera and asked to take a few test pictures with it. If the participant was in the HMD condition, the researcher first gave a tutorial about the features of the prototype task guidance system. The tutorial aimed to familiarize the participant with the system and how to navigate the interface. The tutorial guided the participant in navigating a few cards on the system and the researcher explained different parts of the interface such as the scrollbar and bubble sidebar to the participant. A cheatsheet of voice commands was left with the participants if they were using the voice input. After the introduction, the researcher also gave a tutorial on how to use the power screwdriver. This extra information was provided, as our pre-tests showed participants unfamiliar with tools having trouble with completing the task.

Once the introduction was completed, the participant was left in the room alone to work on the task. If the participant had trouble, they were allowed to go out and request help from the experimenter. A few participants requested help due to difficulty of unscrewing tight screws or not knowing how to use a wrench.

After completing the task, the participant was asked to complete a 40-question questionnaire about their experience and participate in a semi-structured interview of including nine questions. Afterwards, the participant was debriefed by the researcher about the goals of the study and was compensated with $10 USD for participating in the study. The average time for the study was one hour.

## MEASURES

### Objective Measures

In our study, we only collected one objective measure, task completion time, which composed of two parts: the time participant spent on *dissembling* the stepper and the time participant spent on *reassembling* the stepper. We also subtracted the time when the experimenter was asked to help by the participant. Often these were caused by either a software issue (the system crashing) or mechanical issue (a worn out screw could not be removed). The mean completion time for the study was 1796 seconds (29 minutes 56 seconds) with a standard deviation of 468.41 seconds. The fastest participant finished the task in 840 seconds (14 minutes), whereas the slowest participants took 2679 seconds (44 minutes and 39 seconds).

### Subjective Measures

From our questionnaire data, we constructed eight subjective measures. The questionnaire was composed of 40 seven-point Likert-scale questions.

*Usability* – This scale measured the perceived helpfulness of the system. It consisted of eight questions and was highly reliable ($Cronbach's\ \alpha = 0.800$).

*System Usability Scale* – We used a modified version of the System Usability Scale (SUS) [2] to measure the usability of the system. The SUS included 10 questions and was highly reliable ($Cronbach's\ \alpha = 0.804$).

*Goal understanding* – The scale included questions about the user understanding of both the general goal and each step's subgoal in successfully finishing the task. An example question is "I understood how each action fit into my overall goal." The scale consisted of five items and was reliable ($Cronbach's\ \alpha = 0.729$).

*Ease of input* – This two-item scale measures the perceive ease of input for the system. An example question is "it was easy to issue command" ($Cronbach's\ \alpha = 0.742$).

*Clarity of instruction* – This five-item scale measured whether the system displayed the content in a reasonable and useful manner. An example of the question is "The instructions were confusing" ($Cronbach's\ \alpha = 0.810$).

*Multi-tasking* – This is a single question scale asking "I was able to do the task and give commands to the system at the same time."

*System expertise* – This is a single question measuring the participant's perceived expertise using system. This was originally part of a perceive of success scale but was removed due to weak reliability.

*Perceived correctness* – This scale involved a single question: "I was confident that I did the task correctly."

*Demographic information* – We also asked additional background questions on age, gender, occupation, hardware skills, and familiarity with HMDs. All variables were individual items except for hardware skills, which was constructed using two items ($Cronbach's\ \alpha = 0.742$).

## RESULTS

To analyze our data, we used Analysis of covariance (ANCOVA). The covariates were selected through an exploratory analysis of confounding factors such as age, gender, hardware skills and familiarity with HMD. We found that only age and

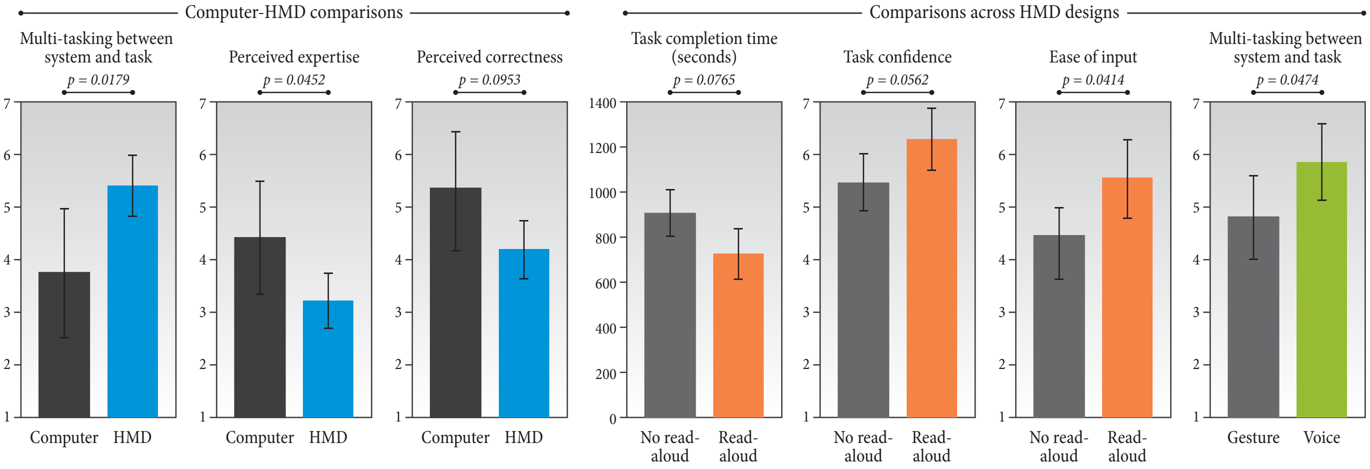


**Figure 6. Comparisons between computer- and HMD-based procedure support (left) and comparisons across HMD designs (right) that show marginal ($p < 0.10$) or significant ($p < 0.05$) differences.**

gender had possible effects on the model, and thus these covariates were subsequently used in all analyses. Our results are summarized in Figures 6 and 7.

*Hypothesis 1 (a)* – The first hypothesis predicted that the glass system would perform better than the computer condition. We did not find evidence supporting this hypothesis. The effect of the type of system used was insignificant ($p = 0.655$) when controlled for both gender and participant's hardware skills. We also ran a Dunnett's test to compare the computer condition against each HMD condition individually and found all comparisons to be insignificant (GGN: $p = 0.874$; GGV: $p = 0.998$; GVN: $p = 0.691$; GVV: $p = 0.935$).

*Hypothesis 1 (b)* – The second part of the first hypothesis predicted that the subjective evaluations of the HMD would be better than the computer. We did not find support for this hypothesis in our analysis of the usability ($p = 0.113$), SUS ($p = 0.772$), goal understanding ($p = 0.463$), ease of input ($p = 0.381$) and clarity of instruction ($p = 0.179$) measures. However, we found that, when asked whether users could multi-task between the system and the task, computer users ($M = 3.714$) rated themselves as significantly less likely to multi-task compared to HMD users ($M = 5.4$), $F(1,33) = 6.206, p = 0.0179$. Participants also rated themselves as an expert of the system significantly more with the computer ($M = 4.429$) than with the HMD ($M = 3.2$), $F(1,33) = 4.332, p = 0.0452$. An interesting trend we observed in our analysis was that computer users ($M = 5.286$) believed that they could finish the task marginally faster compared to HMD users ($M = 4.167$), $F(1,33) = 2.950, p = 0.095$.

*Hypothesis 2 (a)* – We predicted that among the HMD users, those using the voice input would perform better than those using gesture-based input. We did not find any evidence to support this hypothesis.

*Hypothesis 2 (b)* – We hypothesized that participants using the HMD through voice input would evaluate the system more positively than those using the HMD through gesture input. We did not find any support for this hypothesis in our analysis of data from the usability ($p = 0.539$), SUS ($p = 0.965$), goal understanding ($p = 0.828$), ease of input ($p = 0.957$), and clarity of instruction ($p = 0.335$) scales. We found a significant effect on participants' perception on the ease in which the system supported context switching; participants with voice control ($M = 5.875$) rated themselves as more able to multi-task between the system and the task compared to those using gesture control ($M = 4.857$), $F(1,24) = 4.3697 p = 0.0474$.

*Hypothesis 3 (a)* – This hypothesis stated that in users of the HMD system with the text read-aloud capability would perform the task faster than those without. Our data provided only partial support for this hypothesis. While the use of the read-aloud capability did not improve total task completion time, when we split the task into two phases — dissembling and reassembling the stair stepper — we found that the use of the read-aloud capability ($M = 760.429$ sec) marginally improved user performance over not using this capability ($M = 892.25$ sec), $F(1,24) = 3.427, p = 0.0765$.

*Hypothesis 3 (b)* – We predicted that those using the HMD with the read-aloud capability would rate the system more favorably than those without. We did not find any support for this prediction in the data from the usability ($p = 0.790$), SUS ($p = 0.649$), goal understanding ($p = 0.710$), clarity of instruction ($p = 0.603$), and multi-tasking ($p = 0.348$) scales. Surprisingly, we found that users of the system with the read-aloud capability rated the system ($M = 5.357$) better compare to users of the HMD with no read-aloud capability ($M = 4.563$), $F(1,24) = 4.645, p = 0.0414$, regardless of input method. We also found that the use of the read-aloud capability had a marginal effect on participants' perceived success in the task. Participants using the read-aloud feature ($M = 6.07$) rated themselves as more confident in completing the task correctly than those without the system did ($M = 5.563$) $F(1,24) = 4.025, p = 0.0562$.

*Other results* – In addition to the results discussed above, we also found a number of interaction effects in some of our measures. First, we found a significant interaction effect between input and output methods over participants' goal understand-

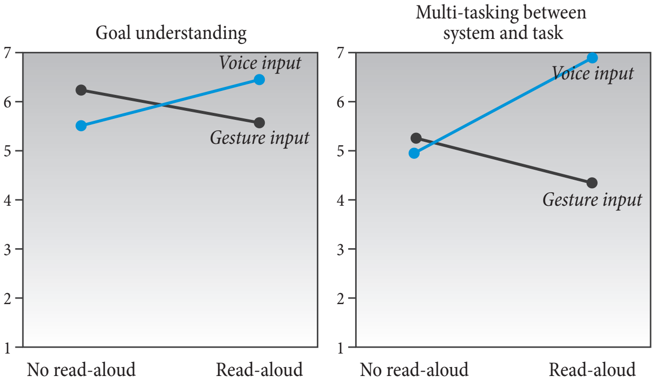


**Figure 7. Interaction effects between HMD input (gesture vs. voice) and HMD output (no read-aloud vs. read-aloud) that were significant ($p < 0.50$) in our data.**

ing ($p = 0.0284$). In particular, among participants using gesture-based control, those using the system with no read-aloud functionality ($M = 6.241$) had poorer goal understanding than those using read-aloud functionality did ($M = 5.569$). On the other hand, participants using voice control rated themselves as having better goal understanding when they used the system without the read-aloud capability ($M = 5.512$) than they did when they used the system with the read-aloud capability ($M = 6.44$). We also found a significant interaction effect between the input and output methods on participants' perception of their ability to multi-task between the system and the task ($p = 0.014$). A post-hoc analysis showed that, voice control has an effect on participant multi-tasking only when the read-aloud capability was used, $F(1, 24) = 10.138$, $p = 0.004$.

## DISCUSSION

Though we did not find significant results in the objective measures to support our hypothesis that our task guidance system performs better than the baseline condition, we did find that our system performs no worse than the computer-based baseline condition. While participants generally believe that they have more expertise and feel more confident in the computer condition, these beliefs did not translate into the computer-based condition outperforming our prototype. Furthermore, the computer-based software, PRIDE view, has been extensively developed and used in the field. Thus, compared to the industry standard, our prototype performed no worse, and the selection of device to be used as a task guidance system will need to be determined by other factors.

Users' perceptions that they were able to navigate the interface and perform the task better when using the HMDs compared to the computer-based condition supports the rationale behind our first hypothesis, as the distance of the computer that provides the instruction will affect task performance due to the need to move around to see the interface. However, the lack of results in the objective measures suggests that there are alternative factors that may have influenced the result. After reviewing the video recordings of the study, we noticed a few possible explanations. One possible explanation might be the closeness of the system to the task space. The computer system was placed directly next to the task and participants did not need to physically move around. We also observed that participants were able to work on the system and look at the screen by orientating their body 90 degrees while their hands remained on the objects in the task space. Therefore, the closeness of the system towards the task space might have affected the results. Another possibility for the lack of objective measure might be the method of displaying instruction. Due to the limited screen size on Google Glass, we were forced to only display one instruction on one card and participants were forced to navigate them slowly. To access the additional content, the user also had to navigate downwards. Furthermore, some participants mentioned that the pictures on the Google Glass were sometimes unhelpful.

> *GGN8185*:"The pictures are useful, but the pictures are sometimes too small"

On the computer interface, there were locations for one or more instructions and often the pictures were directly below the prior instructions. We believe that participants in the computer-based condition were able to collect more information about the task in a single read-through compared to the HMD condition where they had to navigate through multiple cards to get the same amount of information. This might have also been influenced by the fact that instructions were unclear to certain participants.

> *GVV7562*:"some of the instruction were kind of confusing, I have not known at first what they mean by cylinder, I thought that this might be [points to the upper bar] part of it, and so just the wording, I'm not really familiar with tools and anything either. So just that the wording on the slide ..."

The unclear wording combined with unfamiliarity with the names would require the user to navigate to the additional content slide to understand their meaning, forcing users to at least navigate the system one more time compared to the computer-based system which allowed users to see the content on the same page. When we were designing the system, we tried to provide additional content about future steps through the context view. However, as revealed in our interviews, none of the participants actively used the context view nor found any beneficial use for it. This finding might be due to the context view not providing enough information to be useful. While participants did talk about the look-ahead feature in their post-experiment interview, participants often used it to see what the end product would be; this information was not provided by the context view.

> *GVN4704*: "Sometimes it's good to know like the next three... uhm... four steps of the instruction, so that you know this is what we are working towards..."

Similarly, in the HMD condition, we did not observe any effect from method of control on task completion time. We did observe a significant effect on participants' perception of multi-tasking between navigating. In the method of control in the HMD condition, we often observed users having to switch hands holding the tool in order to be able to navigate the interfaces. Participants were also observed to have trouble navigating the interface when there were instructions requiring two hands. For instance, when asking one participant in the Gesture and Voice condition about whether the device ever got in their way, this participant mentioned their difficulty in navigating the interface with both hands occupied.

> *GGV3901*: "the only thing would be like if I need to go [showing the moving forward gesture], if I were like hold up the steps [showing his two hands are occupied], and then told me that I have to change the instructions forwards [one hand move to the moving forward gesture next to the head], and then go back..."

Such instances demonstrate the benefit of a voice control system when involved in a task requiring multiple hands in the task space. However, such a potential benefit did not translate into an improvement in performance. We believe that one possible explanation might be that there is a higher cost to use voice control than swiping forward using the hands, as it may be easier and faster for users to move their hands than to say the command phrase: "forward glass."

However, the interaction effect we observed demonstrated that the effects of voice control in multi-tasking only existed with the read aloud system. A possible explanation for this might be that the cost of reading the text on the screen affects the ability of users to multi-task. Because the Google Glass's screen does not block the field of view of the user, the user must consciously make an effort to look upwards and read the text. Such an action moves the focus of attention away from the task space and onto the Google Glass, and in turn affects users' ability to multi-task between the system and the task. A read aloud system does not require the user to move their visual attention away from the task space and could lead to less shifting of attention. The validity of these claims will need to more thoroughly examined in future work.

Participants utilizing the read aloud feature believed themselves to be more successful at the task than those without. As evidenced by the post-experiment interviews, one possible explanation is that the text produced by the read aloud system acts as a reinforcement for the participants as they work on the task. One participant talked about how he enjoyed listening to the instructions as he worked on the task.

> *GVV7562*: "It's nice hearing it when you try to focus on putting the screws in and everything ..."

As for the significantly better rating on the ease of input for users with the read aloud feature, we could find no clear explanation from the observation of video recording or interview data. One possible explanation might be that participants with the read aloud feature often waited for the read aloud to finish before moving forward and were more careful about aggressive or simultaneous multiple inputs. However, this is only a working idea and would require further testing to confirm.

## LIMITATIONS & FUTURE WORK

Our study has several limitations that might affect the applicability and generalizability of our results. A key limitation was software reliability; a large number of participants experienced bugs that would cause the system to skip certain cards when they were navigating the system or the system crashing at different stages of the task for unknown reasons. The most common issue was a system failure that manifested itself when participants attempted to take photos that eventually led us to exclude this phase of the task from our analysis of the objective measure. While we have taken into account the times and duration of system failures in our analysis, there were additional bugs and issues that did not result in system crashes that were not factored out in our analysis. Furthermore, participants might have reflected their negative user experience resulting from these issues in their responses to the subjective measures.

The complexity of the task also proved to be a challenge for us to control for. Multiple participants experienced problems in unscrewing and screwing parts in because of the wear and tear on the screws caused by the overuse of these parts across trials. In particular, we had to exclude data from one participant because the participant was unable to complete the task due to a broken screw. Over the course of the study, we changed the screws three times to prevent the issue from happening. Additionally, when participants reassembled the equipment, they often fastened parts too loosely or tightly, resulting in a lack of reliability in the tightness of screws across participants. These factors might have affected the validity of the study. We recommend that future studies simplify the experimental task by eliminating screws or other sources of extreme task variability to control for these potential effects.

The complexity of the task also proved to be a challenge for our measurements. While we tried to measure task and ordering errors, participants often employed creative ways to address these errors, as also observed by Zheng et al. [20], followed instructions out of order, or took steps to make the task easier, such as flipping the stair stepper over to access screws more easily. Given the high variability of these actions, we were not able to code for errors, and our resulting objective measure showed extremely high variability.

Our study was exploratory in nature and had the goal of improving our understanding of which factors might be affected by using an HMD system compared to the original computer-based system. Some of our results demonstrate the possibility for further exploration. One of these explorations should be to understand the effects of screen size on task outcomes and how to efficiently present instructions on a small screen to provide additional information. Other future studies should investigate the benefits of navigating an interface with tasks that are less intensive in terms of needing both hands to complete, compared with tasks that require more use of the hands.

## CONCLUSION

As HMDs become more commercially available, there will be an increasing need to better understand their role in job related contexts. Here we present work that is a first step in understanding important factors and benefits in using HMDs as a task guidance system. We developed a prototype task guidance system for Google Glass and conducted a study in which participants used our system to help them disassemble and reassemble exercise stepper equipment. We compared our system with a computer-based task guidance system. While we did not find any significant difference between the computer-based system and our prototype, we found evidence that the HMD might better support the ability to navigate both the task guidance system and the actual task. We also found that a voice control interface provides improved support for multi-tasking between interfaces.

## REFERENCES

1. KM Baird and Woodrow Barfield. 1999. Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality* 4, 4 (1999), 250–259.
2. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
3. Stuart K. Card, Thomas P. Moran, and Allen Newell. 1980. The Keystroke-level Model for User Performance Time with Interactive Systems. *Commun. ACM* 23, 7 (July 1980), 396–410. DOI: http://dx.doi.org/10.1145/358886.358895
4. S. De Buck, F. Maes, J. Ector, J. Bogaert, S. Dymarkowski, H. Heidbuchel, and P. Suetens. 2005. An augmented reality system for patient-specific guidance of cardiac catheter ablation procedures. *Medical Imaging, IEEE Transactions on* 24, 11 (Nov 2005), 1512–1524. DOI: http://dx.doi.org/10.1109/TMI.2005.857661
5. PJ Edwards, DJ Hawkes, DLG Hill, D Jewell, R Spink, A Strong, and M Gleeson. 1995. Augmentation of reality using an operating microscope for otolaryngology and neurosurgical guidance. *Computer Aided Surgery* 1, 3 (1995), 172–178.
6. S. Henderson and S. Feiner. 2011. Exploring the Benefits of Augmented Reality Documentation for Maintenance and Repair. *Visualization and Computer Graphics, IEEE Transactions on* 17, 10 (Oct 2011), 1355–1368. DOI: http://dx.doi.org/10.1109/TVCG.2010.245
7. David Kortenkamp, R Peter Bonasso, Debra Schreckenghost, K Michael Dalal, Vandi Verma, and Lui Wang. 2008. A procedure representation language for human spaceflight operations. In *Proceedings of the 9th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS-08)*.
8. KA Latorella. 1996. Investigating interruptions: An example from the flightdeck. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 40. SAGE Publications, 249–253.
9. Charlotte Linde. 1987. *Checklist interruption and resumption: A linguistic study*. Technical Report NASA–CR–177460. The National Aviation and Space Administration, Washington, DC.
10. Irene C Michas and Dianne C Berry. 2000. Learning a procedural task: effectiveness of multimedia presentations. *Applied Cognitive Psychology* 14, 6 (2000), 555–575.
11. Jennifer Ockerman and Amy Pritchett. 2000. A review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics* 4, 3 (2000), 191–212.
12. Jennifer J Ockerman, Lawrence J Najjar, and J Christopher Thompson. 1997. Wearable computers for performance support: Initial feasibility study. *Personal Technologies* 1, 4 (1997), 251–259.
13. Jennifer J Ockerman and Amy R Pritchett. 1998. Preliminary investigation of wearable computers for task guidance in aircraft inspection. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*. IEEE, 33–40.
14. Michael Rosenthal, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D Pisano, Michael Jiroutek, Keith Muller, Henry Fuchs, and others. 2001. Augmented reality guidance for needle biopsies: A randomized, controlled trial in phantoms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*. Springer, 240–248.
15. Emilie M Roth. 1997. Analysis of decision making in nuclear power plant emergencies: An investigation of aided decision making. *Naturalistic decision making* 175 (1997), 182.
16. Philipp M. Scholl, Matthias Wille, and Kristof Van Laerhoven. 2015. Wearables in the Wet Lab: A Laboratory System for Capturing and Guiding Experiments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 589–599. DOI: http://dx.doi.org/10.1145/2750858.2807547
17. Jane Siegel and Malcolm Bauer. 1997. A field usability evaluation of a wearable system. In *Wearable Computers, 1997. Digest of Papers., First International Symposium on*. IEEE, 18–22.
18. Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: Projected Visualizations for Hand Movement Guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 179–188. DOI: http://dx.doi.org/10.1145/2207676.2207702
19. Kimberly A Weaver, Hannes Baumann, Thad Starner, Hendrick Iben, and Michael Lawo. 2010. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1695–1704.
20. Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. Eye-Wearable Technology for Machine Maintenance: Effects of Display Position and Hands-free Operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2125–2134. DOI: http://dx.doi.org/10.1145/2702123.2702305