Comparing Order Picking Assisted by Head-Up Display versus Pick-by-Light with Explicit Pick Confirmation

Xiaolong Wu[1], Malcolm Haynes[1], Yixin Zhang[1], Ziyi Jiang[1]
Zhengyang Shen[1], Anhong Guo[2], Thad Starner[1], Scott Gilliland[1]

[1]Georgia Institute of Technology
Atlanta GA 30332, USA
xwu86, mghaynes, yixinzhang, jzy0706, szy91,
thad, scott.gilliland@gatech.edu

[2]Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
anhongg@cs.cmu.edu

ABSTRACT
Manual order picking is an important part of distribution. Many techniques have been proposed to improve pick efficiency and accuracy. Previous studies compared pick-by-HUD (Head-Up Display) with pick-by-light but without the explicit pick confirmation that is typical in industrial environments. We compare a pick-by-light system designed to emulate deployed systems with a pick-by-HUD system using Google Glass. The pick-by-light system tested 50% slower than pick-by-HUD and required a higher workload. The number of errors committed and picker preference showed no statistically significant difference.

Author Keywords
Order Picking; Wearable Computers; Head-Up Display

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g., HCI): User Interfaces-Evaluation/methodology

INTRODUCTION
Order picking is the task of collecting items from inventory in a warehouse and sorting them for distribution. This task is an integral step in supply chain operations, and it constitutes a major operational cost, accounting for 55% or more of warehousing expenses [5]. Due to its complex and ever changing nature, order picking is primarily done by human labor. Correctly performing this task is vital. An incorrect pick can stop a manufacturing line or result in dissatisfied customers. Therefore, any improvement in order picking speed and accuracy can result in significant economic gain.

In Table 1 we summarize our previous studies in this area, highlighting the results and key remaining questions. In the study evaluating pick-by-paper, pick-by-light, pick-by-HUD, and pick-by-CMD (Cart-Mounted Display) [3], the results were surprising in that pick-by-HUD was clearly superior in both speed and accuracy. However, the pick-by-light system

Study	Result	Key Remaining Question
HUD vs. audio vs. paper chart vs. paper list [6]	HUD » all in speed and preference; accuracy trends higher	How does HUD compare to light which is widely used in industry?
Variations on HUD interface [1]	Speed and accuracy trends higher adding color & symbols	
HMD vs. CMD [2]	Qualitative feedback at automobile factory	
HUD vs. CMD vs. light vs. paper [3]	HUD » light and paper in speed, accuracy, and preference	How does HUD compare to light with confirmation?
transparent vs. opaque Glass [3]	Opaque 3% faster than transparent	
HUD vs. light + button confirmation	HUD » light in speed; preference trends higher; accuracy trends lower	Compare HUD + scales to light + button + scales?

Table 1: Summary of studies to date.

did not follow the industry standard of using buttons for pick confirmation. In commercial settings, pick-by-light systems require order pickers to push a button after they have picked items from a bin. Once pressed, the system turns off the light for the bin. These lights aid order pickers in knowing which bins have not yet been picked. Without pick confirmation, it is likely that error rates for pick-by-light were higher than typical error rates in commercial settings. Here, we compare pick-by-HUD to a pick-by-light system designed to minimize errors in a manner typical of current industrial implementations. Based on previous results, we hypothesize that the pick-by-light method will be significantly slower than pick-by-HUD but will have fewer errors.

IMPLEMENTATION
In previous work, the predominate type of error for the experimental pick-by-light system was skipping bins (failing to pick any parts from a required bin instead of picking too few or too many of a part from that bin). When working close to the shelves and holding the parts to be picked, the pickers often did not see all of the lit pick bins and would skip one by accident. This error is the most egregious since it could

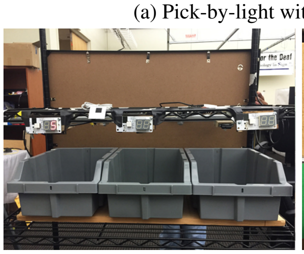
ISWC '15, September 7-11, 2015, Osaka, Japan.

http://dx.doi.org/10.1145/2802083.2808408

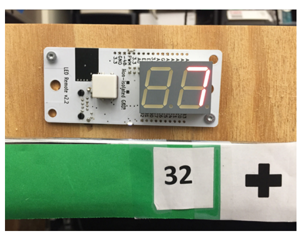
result in several missing parts (from that single bin) and an assembly line halting while the missing parts are retrieved.

We have implemented a pick-by-light system with button-based pick confirmation and error detection (Figure 1). Our pick-by-light system uses LED displays to guide the picker to the correct bin. Each pick bin is associated with an LED display and a button. The display remains off if no items need to be picked from the bin. Otherwise, the display shows the number of items to pick from that bin. After picking the items from each bin, the order picker presses the associated button to confirm the pick and turn off the display. Each order bin (into which the parts are placed) is also instrumented with an LED and button. The LED displays the total number of items that go in the order bin. When the picker presses the button, it updates the display with the number of items left to place in the order bin. To calculate this number, the system sums the number of items left in pick bins whose buttons have not been pressed. This method provides a simple error detection mechanism. Once the remaining number of items is zero, the order bin display turns off and proceeds to the next task.

(a) Pick-by-light with button feedback.



(b) Order bin with button.



(c) Detail view of button board.

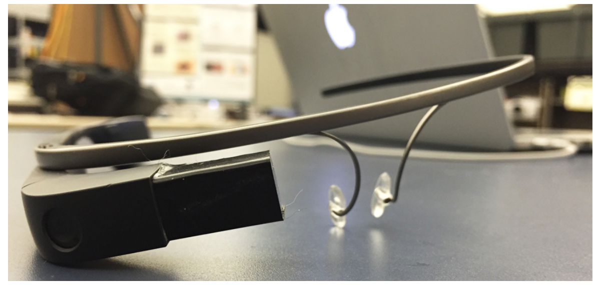
Figure 1: Pick-by-light with button confirmation.

In a pick-by-HUD system, a graphical display guides the order picker to the correct bins. The HUD overlays a graphical depiction of the order requirements in the picker's field of view. Previous studies have shown that a HUD can improve the speed of order picking by 38% compared to pick-by-paper while virtually eliminating errors [6]. Pick-by-HUD has also been compared favorably to pick-by-voice [6].

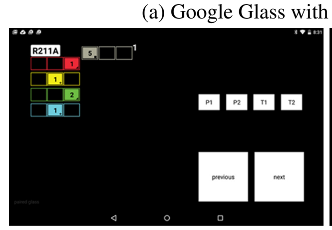
Previous work showed that color coding and symbol coding rows and columns helped improve pick-by-HUD speed and accuracy [2]. Other work examined the performance of transparent and opaque HUDs and found that an opaque display is more efficient for order picking than a transparent display [3]. Our study uses a HUD with an opaque display, as well as color and symbol coded bin representations (Figure 2).

METHODOLOGY

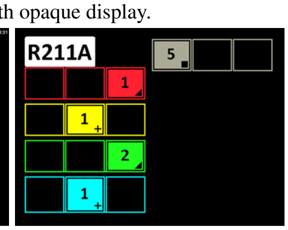
We simulated an order picking environment in our research lab. This environment and apparatus is the same as used in previous order picking studies [3]. The warehouse consists of inventory storage shelves and an order cart. The storage shelves house 24 pick bins. Each bin contains approximately 30 items, and the items are unique for each bin. The pick bins are split evenly between two shelving units. Each shelving unit holds four rows of bins, with each row containing 3 bins. This arrangement results in a simple *shelving unit, row, column* marking system. The rows of the shelves are color coded, top to bottom, with red, yellow, green, and blue. The columns are coded with geometric shapes: square, cross, and triangle, from left to right. This format permits an alternative *shelving unit, color, shape* marking system. Previous research [1] guided the choice of coding. The order cart is a wheeled, mobile cart that contains three order bins. Each order bin has an associated geometric shape, which allows designation by either number or shape.



(a) Google Glass with opaque display.



(b) Android tablet controller.



(c) Glass display.

Figure 2: Pick-by-HUD system.

Task Description

A *task* consists of four to six *subtasks*. A *subtask* requires picking items from a set of pick bins and placing the items in one of the three designated order bins. Completing a subtask consists of a set of picks and a place. A *pick* consists of removing one or more items from a pick bin, and a *place* consists of putting the removed items into an order bin [3]. For each subtask, we randomly assigned up to seven items located in up to five different pick bins on a single shelving unit. We also assigned an order bin for the picked items. An example subtask can be represented as A21:1;2, which translates as go to shelf A, row 2, column 1 to pick 1 item and place it in order bin 2. Leveraging our previous experience, we designed the tasks to be sufficiently complicated so as to induce errors with the hope of having a sufficient number for meaningful comparisons between conditions.

Environment

We recorded each participant with two ceiling mounted cameras. We used three experimenters in the study. One experimenter monitored the study and controlled the tasks displayed in the HUD during the pick-by-HUD evaluations, A second experimenter controlled the order bins. After each subtask, this experimenter recorded the items in the bins, emptied the bins, and replaced the bins on the order cart. A third experimenter controlled the order bin buttons. We used a Wizard-of-Oz technique for the order bin buttons whereby when a user pressed an order bin button, the experimenter manually triggered an update event in the pick-by-light system.

Procedure

We evaluated the pick methods (pick-by-light or pick-by-HUD) with eight participants (two female), all university students, between the ages of 22 and 27 ($M = 23.5$). Six are right-eye dominant and two are left-eye dominant. All participants were first-time order pickers. Each participant was instructed to complete the tasks quickly and accurately.

After receiving instructions, participants completed 10 practice tasks for each pick method (for a total of 20 practice tasks). The number of practice tasks were chosen to extinguish learning effects by the time the test tasks began. After a rest, participants proceeded to the testing session where they completed 10 test tasks for each pick method (20 total). The order of pick methods was counterbalanced using a Latin Square to prevent order effects.

For pick-by-light, the tasks were displayed on the picking bin and order cart LEDs. For pick-by-HUD, the tasks were displayed on a Google Glass controlled by an Android tablet.

After each testing round, participants completed a NASA Task Load Index Survey (NASA-TLX) [4] to measure the workload of the two pick methods. After all testing rounds, participants were asked to answer a survey with regard to overall preference, learnability, comfort, speed and accuracy.

RESULTS

We measured error per pick, error type, task time, task load, and user preferences. For each pick method, we discard the first two of the ten test tasks. This procedure helps further reduce learning effects. Thus, only the last eight tasks from each test session were used for analysis. The eight tasks analyzed were the same for all participants although the order was randomized. This method allows for a within-subject, paired samples comparison between the pick methods for each task and should be a more sensitive test than comparing average task performance. Our method of analysis is consistent with previous studies [3].

Task Completion Time

Task completion time was recorded by the pick systems. The systems start recording when the first subtask is loaded. The task is complete once the last subtask is filled. Since stepping back to check the pick-by-light displays is time consuming, we hypothesized that pick-by-light will take more time to complete. A one-tailed paired samples t-test was used to analyze the data, with the significance level set to $\alpha = 0.05$. On average, pick-by-light ($M = 57.2$ sec, $SD = 13.5$) was over 50% slower than pick-by-HUD ($M = 38.0$ sec, $SD = 7.04$), which passed the test of significance $t(63) = 13.5, p < 0.0001$. The 95% confidence interval for light_time - HUD_time is 16.4 - 22.0 sec/task (see Figure 3).

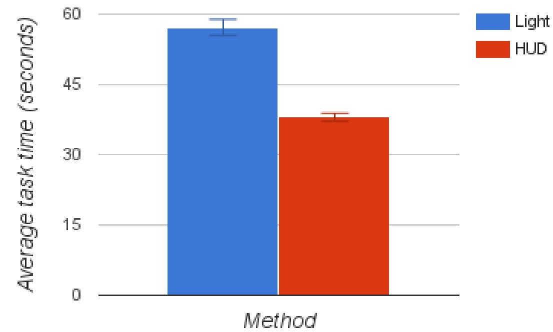


Figure 3: Average task time.

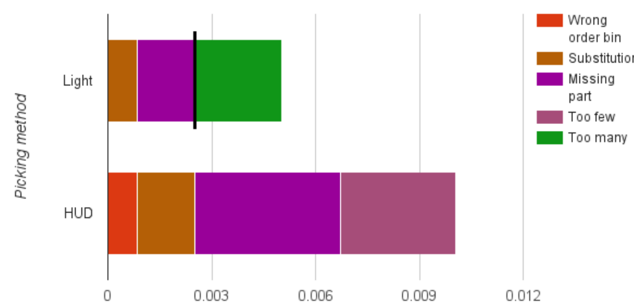


Figure 4: Average error per pick by type.

Error

We found the following types of errors made while picking: wrong order bin (participant put items in a wrong order bin), substitution (participant picked an item from the wrong pick bin), missing part (participant totally missed a pick), too few (participant picked too few items from a bin), and too many (participant picked too many items from a bin). The result is shown in Figure 4. Pick-by-light resulted in an average 0.5% error per pick (6 total) while pick-by-HUD resulted in an average 1.0% error per pick (12 total). There is no statistically significant difference between the two methods ($p = 0.14$, two-tailed), and, even though the experiment is designed to encourage errors, the errors are few enough that it is difficult to ascribe much meaning to them. Even so, putting extra items in an order bin is a relatively benign error, while putting too few items in an order bin is a dangerous error. Future work should focus on these classes of errors and optimize the picking methods to avoid them.

Workload

We used the NASA Task Load Index to measure the workload for each pick method. A TLX score ranges from 0 to 100, with larger numbers corresponding to increased workload. The NASA TLX surveys were analyzed with a one-tail paired samples t-test. Our *a priori* hypothesis was that pick-by-light requires participants to search and correct errors, thereby increasing their frustration and physical workload. Therefore, the results should reflect a heavier overall workload than pick-by-HUD. The results confirmed our hypothesis. Pick-by-light ($M = 58.4, SD = 10.7$) generated a significantly larger workload than pick-by-HUD ($M = 45.7, SD = 9.8$), $t(7) = 2.27, p = 0.029$ (one-tailed).

Preferences
After completing all tests, each participant was asked to rank the two pick methods from best (1) to worst (2) based on overall preference, learnability, comfort, speed, and accuracy. There is no statistically significant difference between the two methods. However, the data trends toward a preference for pick-by-HUD with six of eight participants selecting it over pick-by-light in four of the five categories.

DISCUSSION

Speed
The pick-by-HUD system was significantly faster than pick-by-light. The reasons are twofold. First, our pick-by-HUD system does not require pick confirmation. Users can rapidly move through a task without pausing to press a button. Additionally, participants reported that the HUD provided them with a holistic view of the task. In contrast, the pick-by-light system caused users to focus on individual lights. After picking from a bin, the user often took a step back from the system in order to see all the lights and determine which ones were still on. This step to gain perspective caused an increase in task completion time. Another interesting observation with pick-by-light is that some users counted the items in their hands to ensure they had the right number before placing them into the order bins. Perhaps the order bin pick confirmation became a check on the participant's performance, thereby inducing in some participants a desire to perform well and double check the pick. This counting behavior was not observed in any of the pick-by-HUD tests.

Accuracy
Although there is no statistical difference in the errors generated by the two pick methods, they did result in two different types of errors. Pick-by-light more often resulted in "too many" errors in which participants picked too many items. In contrast, pick-by-HUD primarily resulted in errors of omission in which participants picked too few or wrong items.

We found that the main cause of pick-by-light errors stemmed from a user forgetting to press a button after picking from a bin. This failure is caught at the order bin resulting in the participant thinking they need to pick more items.

The majority of the pick-by-HUD errors stemmed from users picking from the wrong bin or picking too few items from a bin. Perhaps these errors were caused by some pickers taking mental snapshots of a portion of the graphical display and picking based on this snapshot. Such a mental snapshot would require users to remember the color, column, and number of items to pick for each bin. If the participants exceeded the ability of their short term memory to chunk items (which the number of picks in each subtask was designed to do) it could result in incorrect picks. Could a simple check at the order bin, such as a scale to determine whether a reasonable weight of items was dropped, be sufficient to minimize these errors? Or perhaps the number of picks could be limited in each subtask explicitly to avoid these types of errors?

Workload and Preferences
The NASA TLX survey showed that pick-by-light induces a greater workload than pick-by-HUD, likely caused by the process of observing the lights for guidance and error. Additionally, when the system detects an error, it does not tell the user where the error is located. As a result, users have to engage in a search task to find the button they failed to push.

Users preferred pick-by-HUD over pick-by-light. Interestingly, pickers thought pick-by-HUD was more accurate than pick-by-light. The preference for pick-by-HUD may stem from the ability to see the entire subtask at once, giving participants a feeling of control they lack with pick-by-light.

CONCLUSION AND FUTURE WORK
We compared two pick methods, pick-by-light and pick-by-HUD. We implemented industry standard button feedback and error detection on the pick-by-light system. We found that pick-by-HUD was significantly faster than pick-by-light. Pick-by-HUD had a significantly lower workload than pick-by-light, and participants tended to prefer pick-by-HUD over pick-by-light. While we found no statistically significant difference in errors, there was a trend that pick-by-light resulted in fewer errors of omission than pick-by-HUD.

We are currently attempting to improve pick-by-HUD (and pick-by-light) by using scales in the order bins. We hope to build an error detection method that integrates with the systems to inform users of errors as they are picking. Another potential direction is to emulate other picking environments, such as where few parts are distributed to many orders.

REFERENCES
1. Baumann, H., Starner, T., Iben, H., Lewandowski, A., and Zschaler, P. Evaluation of graphical user-interfaces for order picking using head-mounted displays. In *Proc. ICMI '11*, 377–384.
2. Baumann, H., Starner, T., and Zschaler, P. Studying order picking in an operating automobile manufacturing plant. In *Proc. ISWC '12*, 112–113.
3. Guo, A., Wu, X., Shen, Z., Starner, T., Baumann, H., and Gilliland, S. Order picking with head-up displays. *IEEE Computer 48*, 6 (June 2015), 16–24.
4. Hart, S. G., and Staveland, L. E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology 52* (1988), 139–183.
5. Murray, M. Order Picking In The Warehouse. `http://logistics.about.com/od/operationalsupplychain/a/order_pick.htm`.
6. Weaver, K. A., Baumann, H., Starner, T., Iben, H., and Lawo, M. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proc. CHI '10*, 1695–1704.