Order Picking with Head-Up Displays

Anhong Guo, Carnegie Mellon University

Xiaolong Wu, Zhengyang Shen, and Thad Starner, Georgia Tech

Hannes Baumann, Ubimax

Scott Gilliland, Georgia Tech

Experiments suggest that using head-up displays like Google Glass to support parts picking for distribution results in fewer errors than current processes. Making Glass opaque instead of transparent further improves selection efficiency.

Globally, roughly US$1 trillion in goods are distributed from nearly a million warehouse sites each year, and for many businesses such activity represents 20 percent of their logistics costs.[1] Order picking—the process of selecting items from inventory racks with *pick bins* and sorting them into *order bins* for distribution—accounts for about 60 percent of these warehouses' total operational costs.[2]

Current robotic systems lack the dexterity to handle the variety of parts on most pick lines, so the vast majority of western European warehouses still use manual picking,[1] which is costly and time consuming. Although manual methods vary, most warehouses still use paper lists that include each item's location, identifying number, and required amount. Such systems are error prone and can cause significant losses, as in automobile manufacturing, where the wrong part can halt an assembly line. In e-commerce, inventory errors can

compromise order fulfillment, possibly increasing customer dissatisfaction.

Although technologies like parts-to-picker systems that bring parts bins to warehouse workers can facilitate various parts of the picking process,[1] such systems are expensive and relatively rare. Thus, parts bins are typically stationary, and the picker must rely on a paper list or expensive pick-by-light systems that use displays at each bin to indicate which parts to pick.[3]

Wearable computers are becoming more popular with manufacturers as a way to guide pickers, so we constructed an environment for two experiments to compare pick errors and speed. The first looked at paper, pick-by-light, head-up display (HUD), and cart-mounted display (CMD) systems, and the second evaluated a transparent versus opaque version of Google Glass. Our experiments show that HUDs outperform other technologies and that the opaque version of Google Glass might

be more efficient for order picking than the default transparent version.

PICKING METHODS

Picking methods range from the basic paper list to task-guidance systems that are either worn or embedded in the environment. All methods aim to help the picker select the correct parts from the pick bin and deliver them to the correct order bin.

Paper

In the pick-by-paper method, the picker refers to a printout of items to be selected and their locations—one list for each task, which can have multiple orders. The advantages are simplicity and a relatively low implementation cost. On the downside, text-only lists can be difficult to read or interpret when product numbers are long. Often, the first optimization done on a pick line is shortening product numbers to the last three digits (if they are unambiguous).

The list-based picking process forces the picker to perform each step in two stages: read and interpret the list's fine detail and then move a part from one area to another. Often, pickers have the list in one hand while reaching for the parts, which limits the number of parts they can grasp at any one time.[2]

Light

The pick-by-light method is becoming more prevalent in manufacturing warehouses, even though system implementation cost can be as high as US$1,200 per meter. Plant managers report that the virtual elimination of pick errors and the increased picking speed are well worth the investment.[2]

As Figure 1 shows, with pick-by-light, warehouse bins often have small displays and push buttons. When a picker traverses the aisles, the displays illuminate on the pick bins, typically showing the quantity to be picked. Most pick-by-light systems require that pickers press a button to indicate they have picked from the correct bin. In more sophisticated systems, sensors try to detect the picker's reach into each bin. In such systems, the display on a pick bin goes out after each pick is confirmed, and another light illuminates the order bin where the picked parts are to be placed. Pickers tend to dislike the buttons and sensors on pick bins because the sensors do not always correctly detect the reach, which requires more button pushes to fix.

Proximity or weight sensors, or a button press by the picker, indicate when an item is placed in the order bins. Sensor activation or the button press triggers the display of the next task's picks. Although they dislike sensors on pick bins, pickers generally find the order bin sensors valuable.

Cart-mounted display

Pick-by-CMD displays a graphical representation of the picks on the order

FIGURE 1. Pick–by–light method. (a) The pick bin displays the number of parts to pick (1), while (b) the order bin shows where to place (P) the order. (c) Typically, the order bins are mounted on a wheeled order cart so that the pickers can transport the order bins as they walk through the stacks of pick bins in the warehouse.

cart. In some cases, warehouses use a high-resolution LCD display to display picks for an entire shelving unit instead of the pick-by-light systems that indicate each pick bin. Instrumenting the order cart seems more effective and less expensive than instrumenting each shelving unit, but we have seen few such installations.

Head-up display

In this system, the picker wears a HUD that shows the pick charts needed for each shelving unit. Similar to the pick-by-light method, when the picker drops items into the order bin, the HUD displays the next pick chart. Figure 2 shows various technologies suitable for a HUD-based picking system, including an opaque version of Google Glass. Glass is lightweight and self-contained, requires no additional computer, and has sufficient processing power and network connectivity to run order-picking software.

Other methods

Commercial pick-by-voice systems use wearable computers to guide pickers
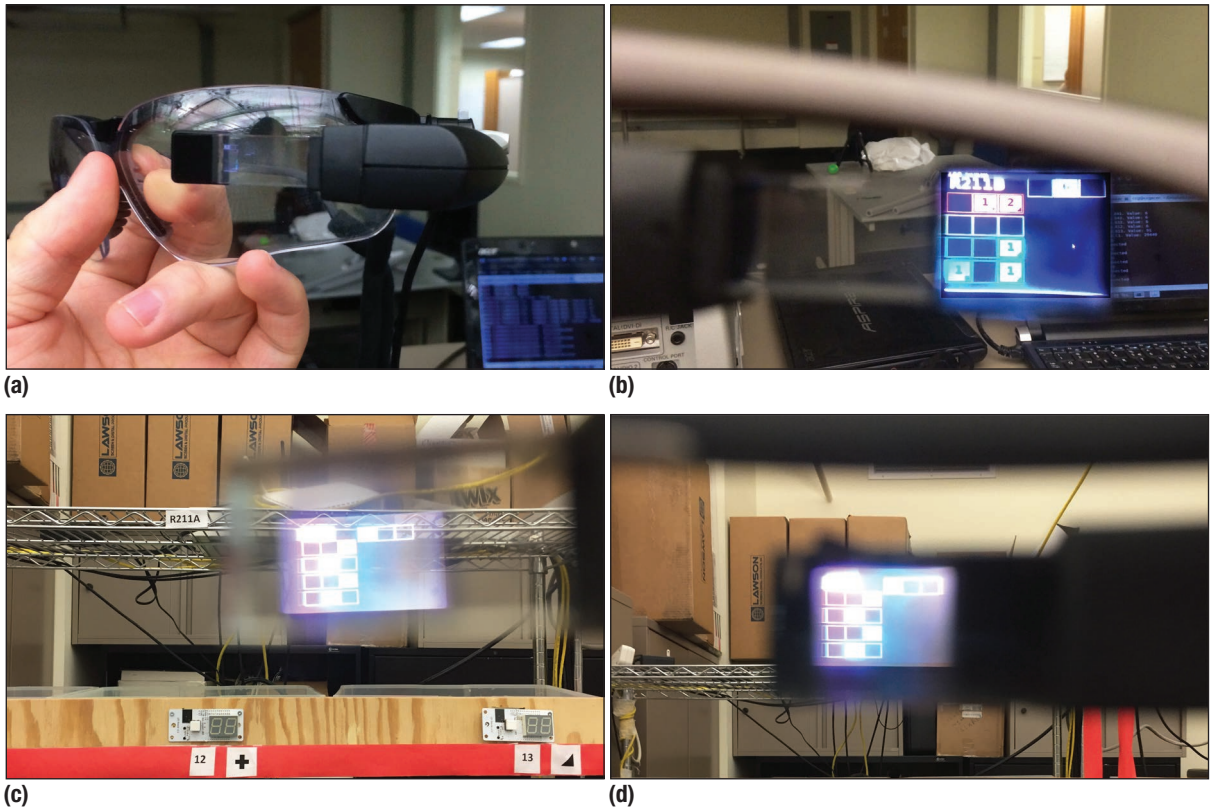
FIGURE 2. Displays used in pick–by–head-up display (HUD) systems. (a) MicroOptical SV–3 HUD mounted on safety goggles. (b) View through the SV–3, (c) view through Google Glass with a transparent display (default), and (d) view through an opaque display.

through their tasks using spoken audio prompts and simple speech recognition. In previous experiments, these systems eliminated many errors but were significantly slower than all other systems, including the standard pick list.[4,5] Similarly, some commercial systems require pickers to confirm each pick by using a handheld scanner to scan a barcode on each pick bin. These systems also eliminate many errors but seem even slower than paper or voice.

Pick-by-vision[6] uses a HUD and motion tracker to overlay 3D graphic tunnels onto the picker's visual field. The tunnels then guide the picker to the right position in the warehouse and highlight the correct pick and order bins with a surrounding frame. Augmented reality systems that use such registered graphics have not yet shown a significant improvement in either accuracy or speed over a paper pick list.

Recent experiments in Germany introduced a pick-by-projector method which guides pickers using video projectors that illuminate pick and order bins. One implementation[5] uses Microsoft Kinect sensors to sense the picker's motions and provide feedback. Preliminary results are promising relative to paper-, voice-, and vision-based picking methods and might eventually prove a viable, albeit more expensive, alternative to the pick-by-HUD system.

EXPERIMENTAL ENVIRONMENT

To evaluate order-picking technologies, we built a dense-picking environment in our research laboratory that uses intelligent storage locations and order batching.[2] Figure 3 shows our experimental environment. Figure 4 shows how the paper- and display-based prompts correspond to the shelves and bins.

Our warehouse environment consisted of 24 pick bins divided between two shelving units, A and B. Each shelving unit has four rows and three columns, and each pick bin contains 20 to 40 items. The order cart, shown in Figure 3b, has three order bins coded with a square, cross, or triangle. The cart's top row holds the paper tasks or the CMD.

We used our environment to conduct two experiments: the first compared the paper, light, HUD, and CMD methods, and the second compared a pick-by-HUD system using transparent and opaque versions of Google Glass. In these experiments, unlike our earlier work,[2,4] we increased task variety, attempting to induce more performance errors for comparison purposes.

We define key terms in our experiments as

› *pick*—one reach into a pick bin and the removal of one or more parts;
› *place*—putting all items currently being carried into an order bin;
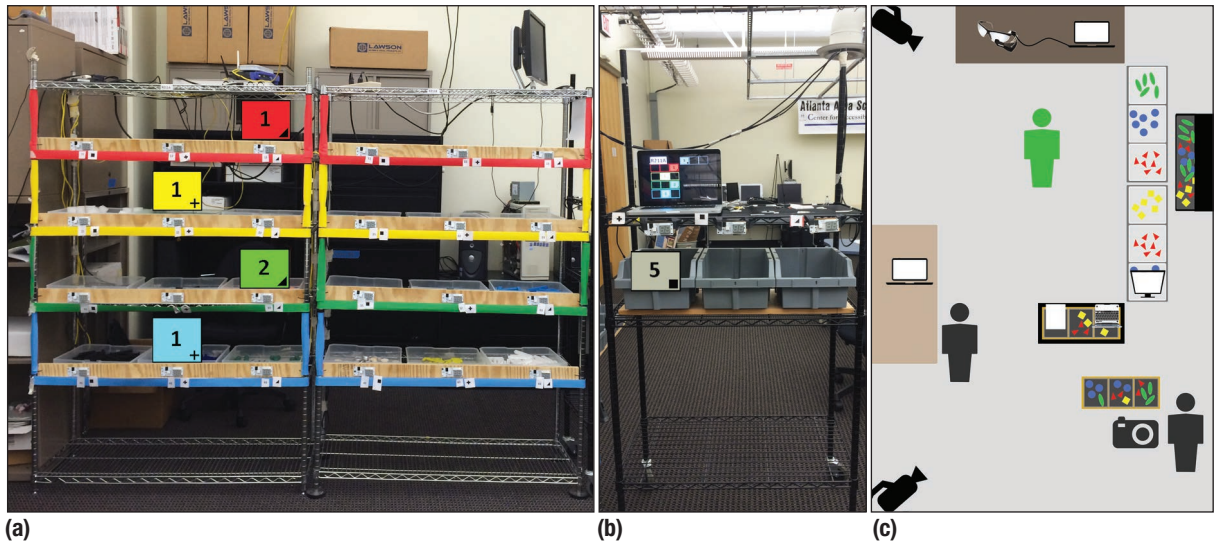
FIGURE 3. Experimental dense–picking environment. (a) The environment consists of 24 pick bins arranged on two shelving units in four rows and three columns per unit. (b) It also contains an order cart with three order bins and a cart–mounted display (CMD) on the top left shelf, which shows a graphical pick chart corresponding to the colored labels in (a). (c) A room diagram shows two video cameras (bottom and top left) to capture the picker's actions, a still camera (bottom right) to capture the picked parts in the order bins for later error analysis, the picker (green figure), and two experimenters (black figures).

> ❯ *task*—a collection of up to six subtasks; and
> ❯ *subtask*—moving one to seven items from one shelving unit to one order bin.

A task consists of up to six subtasks corresponding to three orders: three sets of picks from shelving unit A, placed into three order bins respectively, followed by a set of three sets of picks from shelving unit B, also placed into the three respective order bins. For each subtask, we randomly assigned one to seven items to be picked from one to five pick bins on shelving unit A or B. These items are placed in a single order bin (1, 2, or 3). For example, in Figure 3a, the subtask involves shelving unit A, and 4 of the 12 possible pick bins are involved. The five picked parts are placed into order bin 1 (Figure 3b). Figure 3c shows the layout of the experimental venue.

For our testing, we have simplified the pick list and attempted to make it as efficient as possible. Figure 4a shows a sample. For the first subtask in the sample list, the participant would pick one item from row 1, column 3; one item from row 2, column 2; two

items from row 3, column 3; and one item from row 4, column 2. The participant would place the five items into order bin 1.

Figure 4b shows a pick chart, a graphical representation of the pick list. We use these pick charts for our pick-by-HUD and pick-by-CMD methods. Each pick chart shows the arrangement of the items to be picked for each order bin from each shelving unit. The colors on the pick chart correspond to the colors of the shelf rows shown in Figure 3a. Each row has three bins, and each bin bears a square, cross, or triangle. The pick chart uses these symbols to help cue the picker. In previous research, we showed that such color coding and symbol cues improve pick accuracy.[2] Adjacent colors are arranged so as to be unambiguous to color-blind pickers. Pick charts are significantly more efficient and accurate than pick lists, even when used in paper form, as in Figure 4c.[4]

EXPERIMENT 1: COMPARING FOUR METHODS

In our first experiment, conducted in late 2013 to evaluate the paper, light,

HUD, and CMD picking methods, we enlisted eight participants, ages 22 to 27. All the participants—five males and three females—were novices in order picking. Four were left-eye dominant.

We paid each participant US$20, and the study lasted approximately two hours. The participants were instructed to complete the tasks as quickly and accurately as possible. In both phases, we used a Latin square to counterbalance the order in which the participants performed each picking method.

For pick-by-light, CMD, and HUD systems, a participant could see instructions for only one subtask at a time. After the participant completed each subtask, an experimenter activated the next subtask, emulating the automatic process enabled by a weight or proximity sensor in the order bin. When a participant completed a task, the experimenter replaced the full order bins with empty ones and proceeded to assign the next task. For pick-by-light, we decided not to require the button-push confirmation for picks, since the pickers we interviewed did not like this requirement.
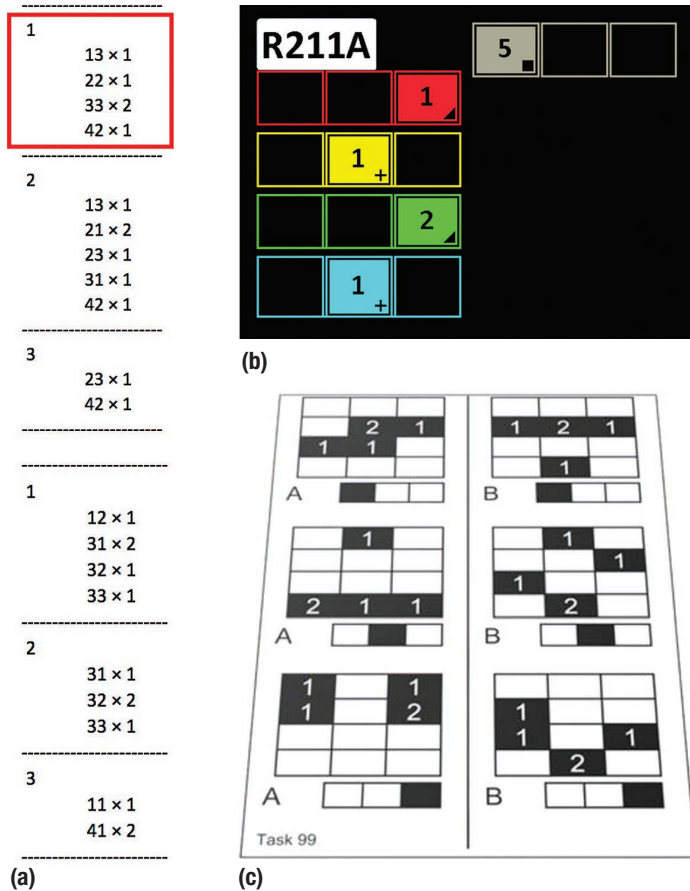
FIGURE 4. Task representations. (a) Sample paper pick list for task 001, subtask 1, on shelving unit A; (b) a graphical pick chart for the same task for the pick–by–CMD and pick–by–HUD methods (overlaid on the pick shelves in Figure 3a); and (c) a paper rendering of the graphical pick chart used in previous experiments.

For pick-by-CMD, we used an LCD-based laptop mounted on the order cart to display the pick chart so that the picker could refer to it with a simple turning of the head while picking. Our pick-by-CMD method was developed based on testing with order pickers at a major automobile manufacturer.[2]

For the HUD system, we used a MicroOptical SV-3 opaque HUD to display pick charts. The display was tethered to a laptop worn in a backpack (Figure 2a) Because Google Glass had not been released at that time, the SV-3 display served as a good proxy, as it has a similar field of view, head weight, and resolution.

For each of the four picking methods, participants performed five practice tasks as part of their training session. Afterward, participants performed 10 test tasks such that each experiment had a total of 20 practice and 40 test tasks.

After using each picking method, participants completed a NASA-TLX (Task Load Index)[7] survey. At the end of the testing phase, they ranked the four methods according to overall preference, ease of learning, comfort, speed, and accuracy.

Errors

To avoid learning-curve effects,[2] our study considered only the last eight tasks from each testing session. We hypothesized that the pick-by-HUD method would have a lower average error per pick and less average task time than the other three approaches. Our experiments were within-subject, and we used one-tailed, paired-sample t-tests, with a significance level of $\alpha = 0.05$.

Picks could have one of three types of errors: *item mistakes*, *wrong number*, and *wrong order bin*. *Item mistakes* had three subcategories:

› *substitution error*—when one part was swapped for another—which could be wrong row, column, or shelving unit;
› *missing-part error*—when the participant omitted a part; and
› *additional-part error*—when the participant placed an unrequested part in an order bin.

Wrong number errors have two subcategories—*too many* or *too few*—occurring when the participant selected too many or too few of the correct parts. *Wrong order-bin errors* occurred when the participant placed the items into the incorrect order bin. Errors of the types *too many* and *additional part* are not severe enough to stop an assembly line and might be discounted or ignored depending on the order picking domain.

Results

Figure 5a shows the methods' average task times. The HUD method's average task time was significantly shorter than using pick-by-light ($p = 0.002$)
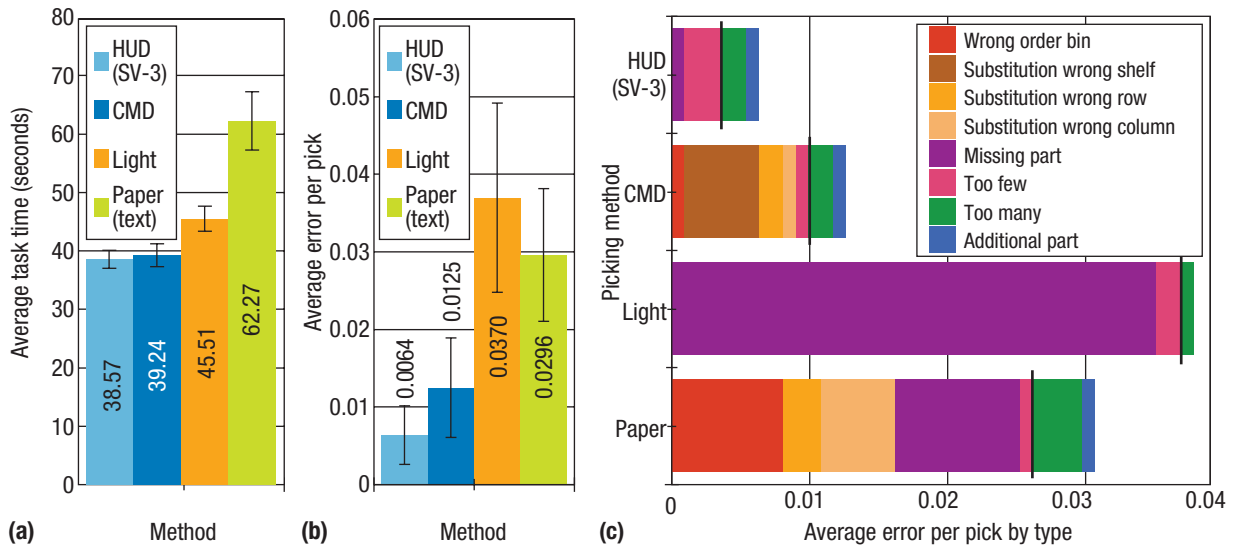
FIGURE 5. Experiment 1 results. (a) Average task time, (b) average error per pick, and (c) average error per pick by type. Errors in (c) are shown from most to least severe, with most severe to the left of the black vertical line.

and paper (p = 0.0001). Figure 5b shows the average error per pick. We counted one error for each error subcategory and used the total number of errors to calculate the method's average error per pick. Using the MicroOptical SV-3 HUD resulted in significantly fewer errors than pick-by-light (p = 0.007) and paper (p = 0.018).[8] Figure 5c shows the errors for all four methods divided by specific error type.

The results supported our hypothesis: the pick-by-HUD method had fewer errors and a shorter task time than either paper or light. Participants also rated the method more favorably, stating that it was much less work. The paper method's poor speed might stem from the need to use one hand to hold the list, and the many errors might be the result of the need to parse the text and remember it while scanning the shelves for the right bin.

HUD method. Our test results show that using pick-by-HUD can virtually eliminate errors and improve speed by approximately 30 percent over paper lists (Figure 5).[8] In addition, the average pick time with the HUD method is a statistically significant speed improvement over pick-by-light. The pick errors with pick-by-HUD,

which were fewer than with any other method, also tend to be less severe. Study participants preferred pick-by-HUD over all other methods tested, and the method had the lowest workload as measured by the NASA TLX.

These results are surprisingly strong. Pick-by-HUD costs considerably less to implement than pick-by-light, making it a promising new tool for order picking. Products such as Ubimax's XPick (www.xpick.de) are bringing the technology to manufacturing environments, but much optimization work remains.

Light method. To our surprise, the pick-by-light method was slower and more error prone than pick-by-HUD. During testing, we saw that pickers were too close to the shelving units to see which other bins were lit, causing them to skip picks and not plan their motions as effectively as they could with a HUD, which offers a task overview.

The low performance is more understandable when considering the participants' view. Pick-by-light users often scan the shelving unit visually from top left to bottom right, which takes time. They also tend to step back frequently to see the entire shelving unit because they cannot keep the

complete context in their heads. Even so, they tend to skip pick bins by accident, as evidenced by the large number of missing part errors in Figure 5c; bins on the shelving unit's periphery were particularly troublesome. Without making the picker press a button to confirm a pick, errors were worse than with a paper list. However, adding a button-push to each pick will slow the process even more.

CMD method. The high performance of the pick-by-CMD method was also a surprise. Speed, accuracy, workload, and preference results approach those of the pick-by-HUD method. However, the errors in the pick-by-CMD method are more critical. As Figure 5c shows, pickers picked from the wrong bin as opposed to not picking the right number of a given part (HUD method).

Perhaps the difference is procedural: pickers turn their heads often between the CMD and the pick bins. A picker who is focused on the cart's display might use peripheral vision to pick and place, resulting in more errors and workload. Regardless of the differences, our results suggest that both pick-by-HUD and pick-by-CMD methods are promising replacements for light and paper methods. Both seem

to reduce errors significantly, have lower setup costs than the pick-by-light method, and provide more flexibility.

EXPERIMENT 2: COMPARING HUD SYSTEMS

In our first experiment, our pick-by-HUD method used MicroOptical's SV-3 display, which performed well but required a tether to a backpack computer. The backpack is both inconvenient and uncomfortably hot.[2] For the second study (transparent versus opaque HUD), we used a self-contained Google Glass device without the backpack or tether, both alone and with a piece of electrical tape added to the Glass display to make it opaque.

Our second experiment used the same environment and most of the same procedure. The goal was to evaluate transparent Glass against an opaque version.

Transparent versus opaque

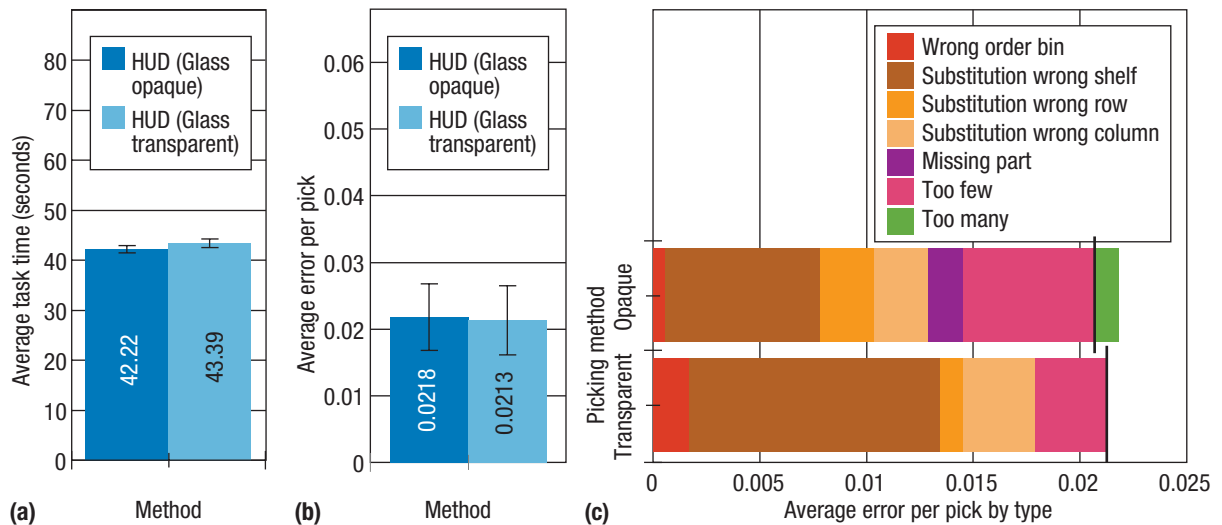Our experiment has some similarity to Robert Laramee and Colin Ware's laboratory study evaluating the effects of transparent and opaque HUDs on a participant's ability to perform a table lookup task on a fixed screen.[9] Participants were stationary, and the task was designed for visual searching, which should be particularly sensitive to the visual interference possible with a transparent HUD. Indeed, their participants performed significantly better using the opaque screen.[9]

We wondered how much visual interference a transparent HUD might cause in order picking, because the user is glancing at the display but is focused more on the environment. We hypothesized that using an opaque display would result in lower average errors per pick and less average task time than a transparent display.

Procedure

We repeated the same testing procedure as in Experiment 1 except that we added 12 participants, 5 females and 7 males. All were ages 21 to 35, new to our studies, and novices at order picking. No one wore regular eyeglasses during the study, since glasses do not fit well with the Glass version we use. All participants who needed vision correction wore contact lenses.

For each of the two picking methods, the participants completed 5 training tasks and then proceeded to 10 test tasks. We again used a Latin square to counterbalance the order of the picking method chosen. In both the transparent and opaque versions, participants saw the same graphical chart (Figures 2c and 2d). To make Glass opaque, we covered the back of the display with black tape.

As in Experiment 1, we used only the last eight tasks from each testing session, randomizing task order for both sessions. We could then conduct a within-subject, paired-sample comparison between the methods for each task, which should be more sensitive than comparing average task performance. We accidentally skipped one task for one participant, which resulted in 95 (12 participants × 8 tasks – 1) task pairs for testing. This number was enough to estimate p values through Monte Carlo permutation methods (2^{95} combinations). Specifically, for each task pair we randomly selected one result for class A, the other for class B, and subtracted B from A. From these 95 differences, we calculated the average. We repeated this calculation process a million times to obtain a distribution of these averages, which should center on zero.

We calculated the actual experimental difference in average task times between the transparent and opaque conditions. Because our hypothesis was that the opaque display would outperform the transparent one, we determined the estimated p value by the percentage of the million trials that shows a difference greater than the experimental difference. We set the significance level at $\alpha = 0.05$ (meaning that only 5 percent of the random groupings show a greater difference). Out of curiosity, we also ran a standard one-tailed, paired t test on the data; the resulting p values were very similar.

Discussion

As Figure 6a shows, the average task time using the opaque display is less than when using the transparent one. The difference passes the significance test ($p = 0.0416$, 99 percent confidence interval [0.0410, 0.0421]). As shown in Figure 6b, the average error per pick for the two methods is close. Figure 6c

> THROUGH REPEATED TESTING WITH ORDER PICKING, HUD OPTICAL PROPERTIES MIGHT BE OPTIMIZED FOR SIMILAR INDUSTRIAL TASKS.

FIGURE 6. Results of comparing opaque and transparent Google Glass versions for order picking. (a) Average task time, (b) average error per pick, and (c) error per pick by type.

shows the errors divided into specific error types.

Although we were correct in our hypothesis that the opaque HUD results in faster picking than the transparent display, the difference is less than 3 percent—much lower than observed in Laramee and Ware's study with a stationary user.[9] Perhaps the difference is procedural; order picking alternates rapidly between virtual and physical worlds, while Laramee's stationary user dealt mostly with the virtual realm, which would maximize any effect from visual interference with the physical world.

A deeper investigation into the data shows a clear learning effect, which surprised us until we realized that we had given participants only 10 practice tasks instead of the 20 practice tasks for participants in the first experiment. This oversight might also explain the higher error rate: 2 percent with Glass overall relative to 0.6 percent with SV-3.

Our experimental design seems to have sufficient sensitivity to help select among picking system variations, enabling several lines for continuing work. Our picking environment is closely modeled after those we observed in the automobile industry, but our results should generalize to similar environments. Indeed, DHL (a major shipping firm), Ricoh, and Ubimax have recently seen a 25 percent improvement in picking performance with smart glasses. However, for pick environments that differ significantly from the one we tested, we could easily modify the experimental environment and retest.

Order picking might prove to be a good reference task for optimizing HUD characteristics for industrial environments, given that it requires constant movement, interaction with the physical world, and quick glances to a pick list or chart. Through repeated testing with order picking, the optical properties of HUDs (for example, brightness, contrast, field of view, color depth, focus, eye box, and bi-ocular versus monocular presentation) might be optimized for similar industrial tasks. A comparison of Glass and the SV-3 is a case in point. Although the two have a similar field of view, Glass is centered above the user's field of vision on the right eye, while the SV-3 is adjustable, and the view is slightly below the participant's line of sight. The SV-3 also has an adjustable focus and is often worn on the dominant eye, left or right. Is the higher error rate observed with Glass versus the SV-3 because of placement as opposed to the learning effect we hypothesized? A relatively quick study could determine the truth. Glass is becoming a popular experimental platform in industry, making it a logical choice for continued testing.

Safety is another important concern. Perhaps a HUD-based picking interface could in fact be safer than current paper lists because the HUD keeps pickers' hands free and is less distracting. If so, comparing the use of new HUD models to standard paper-based pick lists might provide a measure of safety assurance—much as tests in the automotive industry evaluate the safety of new dashboard interfaces by comparing the degree of distraction between using the interface and changing a radio channel.

The poor performance of the pick-by-light method also suggests an area for further study. How will including pick-confirming sensors or a button push affect accuracy and speed relative to the pick-by-HUD method? Adding scales under each order bin could reduce errors for all methods but might be particularly effective against missing parts or picking too few parts of a given type. If including both sensors and scales could virtually eliminate these errors, will pick-by-HUD

ABOUT THE AUTHORS

ANHONG GUO is a doctoral student at Carnegie Mellon University's Human–Computer Interaction (HCI) Institute. His research interests include wearable computing, crowdsourcing, and accessibility. Guo received an MS in HCI from Georgia Tech. Contact him at anhongg@cs.cmu.edu.

XIAOLONG WU is a master's student in HCI at Georgia Tech's School of Interactive Computing. His research interests include HCI, wearable computing, and user experience research. Wu received BEng in industrial design from Shanghai Jiaotong University, China. Contact him at xwu86@gatech.edu.

ZHENGYANG SHEN is a master's student in HCI at Georgia Tech's School of Interactive Computing. His research interests include HCI, information visualization, and ubiquitous computing. Shen received a BEng in electronic information science and technology from Tsinghua University, China. Contact him at szy91@gatech.edu.

THAD STARNER is a professor of computing at Georgia Tech. His research interests include pattern discovery, wearable computing, and sign language. Starner received a PhD in wearable computing from MIT. He is a member of IEEE and ACM. Contact him at thad@gatech.edu.

HANNES BAUMANN is a project leader and software developer at Ubimax in Germany. While performing the work described in this article, he was a doctoral student in the Wearable Computing group of the Technologie-Zentrum Informatik und Informationstechnik, Universität Bremen, Germany. His research interests include HCI, wearable computing, and image processing. Baumann received a PhD in computer science from Universität Bremen. He is a member of the German Informatics Society (GI). Contact him at hannes.baumann@ubimax.de.

SCOTT GILLILAND is a research scientist at the Georgia Tech Interactive Media Technology Center. His research interests include rapid prototyping and embedded systems. Gilliland received an MS in computer science from Georgia Tech. Contact him at scott.gilliland@gatech.edu.

still be faster than pick-by-light and, if so, by how much? Answering these questions is just one of many directions for future research and could move industry more quickly toward assistance in performing a tedious and error-prone process. ⬛

REFERENCES
1. R. De Koster, T. Le-Duc, and K.J. Roodbergen, "Design and Control of Warehouse Order Picking: A Literature Review," *European J. Operational Research,* vol. 182, no. 2, 2007, pp. 481–501.
2. H. Baumann, "Order Picking Supported by Mobile Computing," PhD dissertation, Univ. of Bremen, 2013.
3. M. De Koster, "Warehouse Assessment in a Single Tour," *Warehousing in the Global Supply Chain*, M. Lahmar, ed., Springer, 2012, pp. 457–473.
4. K.A. Weaver et al., "An Empirical Task Analysis of Warehouse Order Picking Using Head-Mounted Displays," *Proc. ACM Conf. Computer-Human Interface* (CHI 10), 2010, pp. 1695–1704.
5. J. Herter, "Augmented Reality Supported Order Picking Using Projected User Interfaces," bachelor's thesis 131, Univ. of Stuttgart, 2014.
6. B. Schwerdtfeger et al., "Pick-by-Vision: There Is Something to Pick at the End of the Augmented Tunnel," *Virtual Reality*, vol. 15, 2011, pp. 213–223.
7. S.G. Hart and L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psychology*, vol. 52, 1988, pp. 139–183.
8. A. Guo et al., "A Comparison of Order Picking Assisted by Head-Up Display (HUD), Cart-Mounted Display (CMD), Light, and Paper Pick List," *Proc. IEEE Int'l Symp. Wearable Computers* (ISWC 14), 2014, pp. 71–78.
9. R.S. Laramee and C. Ware, "Rivalry and Interference with a Head-Mounted Display," *ACM Trans. Computer-Human Interaction*, vol. 9, no. 3, 2002, pp. 238–251.