A Comparison of Order Picking Assisted by Head-Up Display (HUD), Cart-Mounted Display (CMD), Light, and Paper Pick List

Anhong Guo[1], Shashank Raghu[1], Xuwen Xie[1], Saad Ismail[1], Xiaohui Luo[1], Joseph Simoneau[1]
Scott Gilliland[1], Hannes Baumann[2], Caleb Southern[1], Thad Starner[1]

[1]Georgia Institute of Technology
85 Fifth Street NW
Atlanta GA 30332, USA
{guoanhong, sraghu, simoneau, scott.gilliland, caleb.southern, thad}@gatech.edu, {mdjxiexuwen, saad.ismailm, lxhhuihui}@gmail.com

[2]Ubimax GmbH
Wachtstrasse 17-24
28195 Bremen, Germany
hannes.baumann@ubimax.de

ABSTRACT
Wearable and contextually aware technologies have great applicability in task guidance systems. Order picking is the task of collecting items from inventory in a warehouse and sorting them for distribution; this process accounts for about 60% of the total operational costs of these warehouses. Current practice in industry includes paper pick lists and pick-by-light systems. We evaluated order picking assisted by four approaches: head-up display (HUD); cart-mounted display (CMD); pick-by-light; and paper pick list. We report accuracy, error types, task time, subjective task load and user preferences for all four approaches. The findings suggest that pick-by-HUD and pick-by-CMD are superior on all metrics to the current practices of pick-by-paper and pick-by-light.

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

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

INTRODUCTION AND RELATED WORK
Order picking is the process of collecting items from a warehouse inventory and sorting them into orders for distribution. Each year, warehouses throughout the world distribute approximately one trillion dollars in goods from nearly a million warehouses [5]. Order picking is one of the main activities performed in these warehouses, accounting for about 60% of the total operational costs [1].

ISWC'14, September 13 - 17, 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2634317.2634321

Many efforts have been made to improve the speed and accuracy of order picking systems, including the research and development of fully automated picking solutions and parts-to-picker systems that bring bins of parts to warehouse workers [5]. The most common systems, however, are much simplier and less expensive picker-to-parts systems that require a picker to go to stationary part bins [4]. In picker-to-parts systems, it is estimated that 50% of a picker's time is spent traveling from one storage rack to the next, and 35% of time locating and picking from the correct bin [9]. Various strategies are used to reduce travel time, including picking multiple orders at once (batching), picking only specific parts (zoning), and finding more optimal bin placements and picker routing.

To reduce the time needed to pick items once the picker is at the correct storage rack, several task guidance systems are commonly used: text-only paper pick lists (pick-by-paper), illuminated bin indicators (pick-by-light), and audio-assisted picking (pick-by-voice). In previous work, the use of a head-up display (HUD) showed significant picking speed gains as compared to a traditional text-based pick list, a paper-based graphical pick chart, and a mobile pick-by-voice system [10]. Another study showed how to improve pick speed and accuracy with pick-by-HUD by associating colors with shelves and shapes with bins [7]. Although pick-by-light has been compared qualitatively to pick-by-HUD in the field in a commercial facility [3], pick-by-HUD has not been quantitatively evaluated against pick-by-light in a controlled study. The same field study introduced the idea of using a display mounted on the picker's mobile cart to show the needed picking chart, but no quantitative evaluation was performed.

Our main contributions include:

- a laboratory evaluation comparing the speed, accuracy, error types, workload, and preference of pick-by-HUD to pick-by-paper, pick-by-light, and pick-by-CMD; and
- the finding that pick-by-HUD offers superior performance to pick-by-light and pick-by-paper, with potentially much lower cost than pick-by-light.

Furthermore, post-hoc analysis suggests that pick-by-CMD also provides superior performance in all metrics compared to pick-by-paper and pick-by-light but does not prevent as many critical pick errors as pick-by-HUD.

Pick-by-Paper

A paper pick list specifies the location of each type of item, the number of items to be picked, and the sequence in which the items will be picked to guide the picker. This method is the simplest and cheapest system to implement, and is the most commonly used [1], especially in small-scale warehouses.

Text-only paper pick lists can be cumbersome and difficult to read or interpret. They tend to be error-prone as they require the picker to manage big-picture tasks and workload while simultaneously interpreting the specific information from the pick list [1]. Due to their widespread use and evaluation, we have included them in this study as a baseline.

Pick-by-Light

In these systems, sometimes called pick-to-light systems, the warehouse bins are wired with small displays and sensors. As a picker traverses the aisles, the displays light, typically showing the quantity of item to be picked. After the part or parts have been picked, another light will illuminate on the order bin where the part should be placed. In some systems, pickers must press a button on the shelving unit to indicate they have picked from the correct bin. In other systems, a laser scanner is mounted above the shelf and detects the picker's reach into each bin. In practice, however, these scanners are often triggered incorrectly. For example, as a picker reaches into a bottom bin, his head may cross the laser beam first, causing the system to believe that the incorrect item was picked. This type of error might cause the picker to need to hit two buttons to correct the system: the first for indicating the "incorrect" pick was actually not incorrect, and the second to indicate that the correct pick was performed. In interviews with pickers in automobile plants, these errors are considered a major irritant [1]. However, proximity sensors in the order bins, where the parts are placed, are considered very useful as they automatically advance the pick guidance system to the next task. Based on these pickers' recommendations, we decided to simulate the proximity sensor-based detection of places but not the laser-scanner-based detection of picks.

Pick-by-CMD

The pick-by-CMD (Cart-Mounted Display) method emerged from previous research, which showed that a graphical version of the paper pick list performed significantly better than the traditional pick list [1]. Here, we transfer this graphical layout to a display mounted on the picking cart (that holds the order bins).

Pick-by-HUD

A Head-Up Display (HUD) or Head-Mounted Display (HMD) is worn by the picker. The display shows the same graphical layout of the shelving unit as in a CMD system. Weaver et al. showed that using a HUD could improve the speed of order picking by 38% over paper pick lists and virtually eliminate errors [10]. Pick-by-HUD has also been compared favorably to other systems, such as pick-by-voice [7, 10]. The performance of pick-by-HUD has not previously been compared with pick-by-light in the laboratory.

Augmented Reality technology has been tested for order-picking systems, but the additional hardware and set-up time needed to register the graphics does not seem to outweigh the benefit over a 2-D graphical representation [8].

In the types of visual displays used in pick-by-HUD and pick-by-CMD systems, color-coding and symbol-coding rows and columns helped improve speed and accuracy. The addition of part images or descriptions slowed the user by giving them too much to look at for each pick [2].

EXPERIMENTAL DESIGN

Warehouse Layout

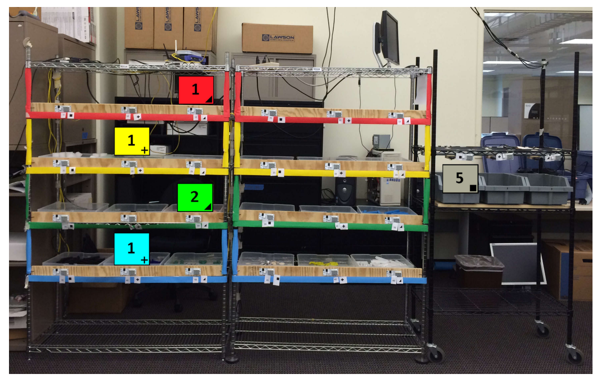


Figure 1: Experimental setup, including 24 pick bins (on two shelving units with four rows and three columns each) and three order bins on the right. An example pick list is annotated with superimposed labels.

We set up a simulated order picking space in our research lab (Figure 1). In order to focus on the picking of the items, we employed a high picking density and minimized the travel time between picking locations. Our picking space would be consistent with a real-world picking scenario that utilized intelligent storage locations and intelligent batching of orders.

The system contains two shelving units, R211A and R211B. For each shelving unit, there are four rows and three columns of pick bins, giving 24 ($2 \times 4 \times 3$) positions in total on both shelving units.

The rows are color coded as red, yellow, green, and blue (top to bottom). Red-green and blue-yellow color blindness are taken into consideration by separating those two color pairs. The columns are coded as symbols: square, cross and triangle. We chose color and symbol coding for the rows and columns based on the previous research [2].

In each position on a shelf, there is a pick bin containing 20 to 40 items of each item type. With this number of items in the bins, there are enough items to prevent any bin from running out.

The other part of the system is the order cart. The top row is used to hold past or future paper tasks or a computer display. The middle row has three order bins fixed together to make a set (Order Bin 1, 2, and 3). The order bins are also symbol coded as square, cross and triangle. The cart also has wheels, so the users were free to move it around when conducting the tasks (although none did).

Task Description

A *pick* is defined as reaching into a pick bin and removing one or more parts. A *place* is defined as putting all of the items currently being carried into an order bin.

A *subtask* is comprised of moving multiple items (between one and seven) from one shelving unit to one order bin. For each subtask, we randomly assigned one to seven items chosen from one to five pick bins located on a single shelving unit (A or B), and destined for a single order bin (1, 2, or 3). Figure 2c shows the representation of a subtask on a screen.

A task is comprised of up to six sub-tasks (2 shelving units $\rightarrow$ 3 order bins), or differently said a batch of three orders. A *method* is one of the picking technologies: pick-by-paper, pick-by-light, pick-by-CMD, pick-by-HUD. For each picking method, participants performed 5 practice tasks and 10 test tasks. Thus, each participant performed a total of 20 practice tasks and 40 test tasks.

In previous work, the error rates observed were extremely low (e.g. 0.4% to 1.9% [10]), and thus not indicative of potential error rates in real-world picking tasks. We believe this is due to unrealistically repetitive tasks. To better simulate the potential for errors in complex real-world picking tasks, we deleted subtasks from some of the tasks in order to introduce variation. In the five practice tasks, we deleted one subtask from two of the tasks. In the ten test tasks, we deleted one subtask from two of the tasks and two subtasks from one of the tasks.

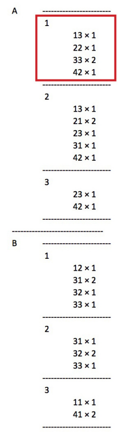
Picking Methods

For all four picking methods below, we use the same pick list example of Practice Session A, Task 0001 (visualized in Figure 1). For the first subtask in this example, the participant is asked to pick 1 item from row 1 column 3, 1 item from row 2 column 2, 2 items from row 3 column 3, and 1 item from row 4 column 2. The five items should be placed into order bin 1.

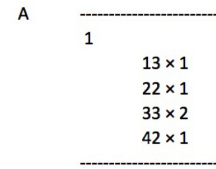
Pick-by-Paper

A paper pick list (Figure 2a) specifies the location of each type of item, the number of items to be picked, and sequence in which the items will be picked. Each paper pick list represents an entire task. The task set this task belongs to is printed on the bottom right corner, and the task number within this task set is printed on the bottom left corner.

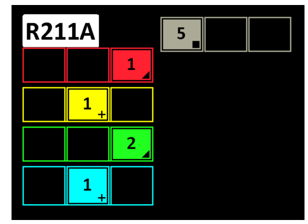
Different shelving units are divided by long dashed lines. Different subtasks are divided by short dashed lines. For each subtask, the positions and items to pick from the positions are specified in lines such as AB$\times$N, where A is the row number, B is the column number, and N is the number of items to pick from that location.



(a) Paper pick list; the red box around practice A, task 0001, subtask 1 added for emphasis.



(b) Practice A, task 0001, subtask 1.



(c) Display for pick-by-CMD and pick-by-HUD.

Figure 2: Different representations of a sample pick list.

Pick-by-Light

The pick and order bins are instrumented with circuit boards, with two-digit 7 segment LEDs and push buttons. For each task, the pick bins displays light up at the corresponding location with numbers on it (Figure 3a), indicating where to pick and how many items to pick. At the same time, one of the three order bins LEDs will light up with a letter P (Figure 3b), indicating which one is the right order bin.

Pick-by-CMD

A graphical display (Figure 2c) is used to represent the pick tasks, which is presented on a cart-mounted display to guide the picker (Figure 4). On the left side, a grid is displayed to represent the layout of the shelving unit, with the shelving unit number on the top left corner, shelving unit rows color-coded, and columns symbol-coded. The bins to pick from are represented visually, along with the number of parts to grab from that particular bin.

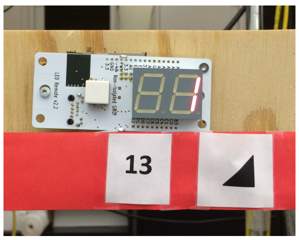
On the right side, another grid represents the layout of the order bins, with the bins symbol-coded. The order bins to put are also represented visually, along with the number of parts to put in that particular bin.

Pick-by-HUD

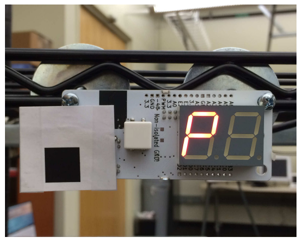
A Head-Up Display (HUD) or Head-Mounted Display (HMD) is used to assist order picking, in that the system guides the picker to each item using shelving unit/bin numbers overlaid on the user's visual field as they transverse the pick area. The user interface is the same as the pick-by-CMD guidance system (Figure 2c).

Environment

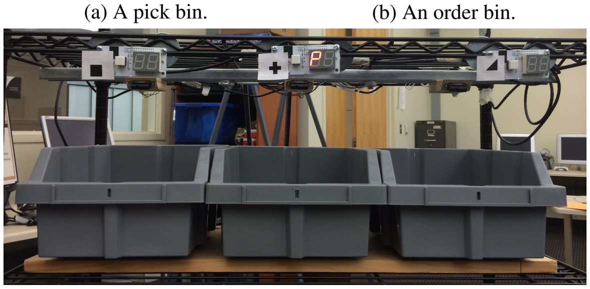
The user study environment is set up as in Figure 5. We recorded the entire study procedure with a front and a back video camera. The study procedure requires at least two experimenters. One experimenter is in charge of monitoring the study. When a subtask is finished, the experimenter clicks the mouse to move forward to the next subtask. Because we are measuring the systems themselves, we decided to use Wizard-of-Oz methods, rather than automate the task using proximity sensors that detect places in the order bins. The second experimenter is in charge of the order bins. After each task, the experimenter takes the order bins off the cart, photographs the order bins, and replaces them with an empty set. Then the participant can continue to the next task. One of the experimenters is in charge of replacing the used items in order bins, without interfering with the participant's task. This experimenter ensures there are at least 20 items in each pick bin, in order to maintain a consistent level of difficulty.



(a) A pick bin.



(b) An order bin.



(c) Pick cart with pick-by-light system.

Figure 3: Displays used for the pick-by-light system.

Figure 4: Pick cart with CMD.

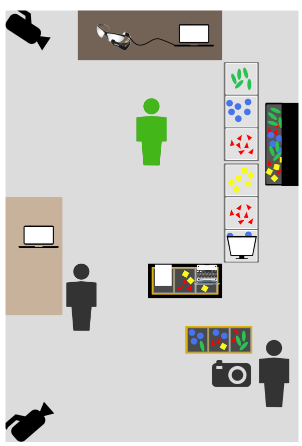


Figure 5: User study setup.

Procedure

We evaluated the systems with eight participants (three female), all students, with an age range from 22 to 27. All eight participants are right-handed. Four are left-eye dominant and four are right-eye dominant. All eight participants were novices at all four picking methods. Participants were compensated with U.S. $ 20 for the study and were instructed to complete the tasks as quickly and as accurately as they could.

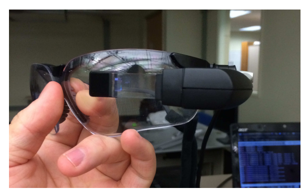
The participants first completed 5 training sessions for each picking method, and then proceeded to complete 10 test sessions for each method. The order in which the participants performed each picking method in both training and testing phases were counterbalanced using a Latin Square.

For pick-by-paper, paper pick lists were printed on US letter paper and placed on their backs on the cart in a pile. After a participant finished a task, he or she placed the finished paper list in another pile. At the same time, an experimenter replaced the order bins. The participant then picked up the next paper pick list and proceeds to the next task.

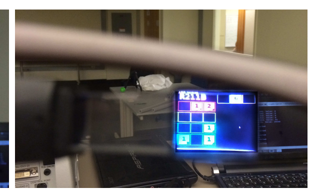
For pick-by-light, instructions for a subtask are presented on the LEDs on the shelving units and on the cart.

For pick-by-CMD, the graphical display is presented on a 13.3 inch Macbook Pro with Retina Display on the cart.

For pick-by-HUD, the graphical display is presented on MicroOptical SV-3, connected to a Acer laptop. The display is worn on a pair of glasses (Figure 6). The participant wears a backpack with the computer inside to power and drive the display.



(a) SV-3 with goggles.



(b) User's view into SV-3.

Figure 6: Displays used for the pick-by-HUD system.

For pick-by-light, CMD, and HUD systems, the participant is able to see instructions for only one subtask at a time. After each subtask is complete, an experimenter clicks the mouse to proceed to the next subtask. After the participant finishes each task, an experimenter replaces the order bins. After the replaced order bins are in place, an experimenter will click to mouse to proceed to the next task.

After each picking method, participants complete a NASA Task Load Index Survey (NASA-TLX) [6] to measure the task load of the systems. At the end of the testing phase, participants are asked to rank the methods from best (1) to worst (4) based on overall preferences, learnability, comfort, speed and accuracy.

RESULTS

We report average error per pick, error types, average task time, task load and user preferences for all four approaches. Only the last 8 tasks from each testing session are used. Our *a priori* hypotheses were that pick-by-HUD has lower average error per pick, less average task time, less task load, and

Error Types	
Item Mistake	Substitution
	Missing Part
	Additional Part
Wrong Number	Too Few
	Too Many
Wrong Order Bin	

Table 1: Error types; each row is considered one error.

higher rankings than the three other approaches. For all the t-tests we ran, the significance level is set to $\alpha = 0.05$.

Average Error Per Pick

We used the images and videos to determine number and type of errors (Table 1). We compared the participant images with standard task images to analyze the pick errors, and examined the videos to determine specific error types. There are three error types: 1) *item mistakes*; 2) *wrong number*; and 3) *wrong order bin*.

For *item mistakes*, there are three sub-categories. *Substitutions* occur when one part was swapped for another part, including three variations: *wrong row*, *wrong column*, and *wrong shelving unit*. *Missing part* means a part was omitted. *Additional part* means an unrequested part was placed in an order bin.

For *wrong number* errors, there are two sub-categories. *Too many error* means too many items of the right part were picked, and *too few error* means too few items of the right part were picked. If the participant did not pick any, it is classified as a missing part error.

A *wrong order bin* error occurs when the participant placed the items into the incorrect order bin.

We counted one error for each row in the error types table (Table 1). The total number of errors are used to calculate the average error per pick metric (Figure 7). A one-tailed paired samples t-test was used to compare participant's average error per pick for all four picking methods. The HUD ($M = 0.006, SD = 0.011$) resulted in significantly fewer errors than light ($M = 0.037, SD = 0.035$), $t(7) = -3.272, p = 0.007$ (one-tailed), and paper ($M = 0.030, SD = 0.024$), $t(7) = -2.608, p = 0.018$ (one-tailed). Figure 8 shows the errors for all four methods divided into specific error types. Errors on the left are the most severe, and errors on the right are the least severe. More number errors and additional parts errors are not severe enough to stop an assembly line; all other errors (shown to the left of the black separator line) could be sufficiently severe to stop an assembly line. Figure 9 shows the positions where errors happened for all four methods.

Average Task Time

We examined the videos to determine the time for each task. For pick-by-paper, start time was defined as when the participant picked up the paper task list. The start time for the pick-by-light was defined as when the LEDs are lit for the first shelf-order combination. For pick-by-HUD and CMD, the start time was defined as when the first shelf-order combination was displayed. For all methods, the end time was determined by when the last item was placed into the order bin.

The average task time for each of the picking methods are presented in Figure 10. A one-tailed paired samples t-test

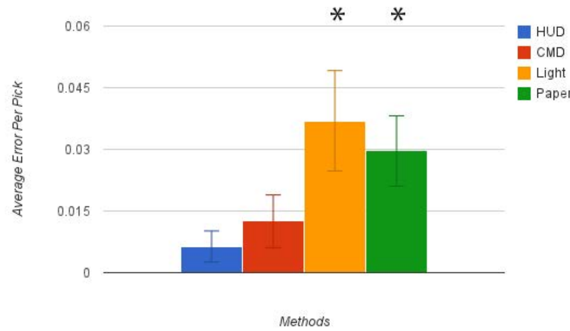


Figure 7: Average error per pick. Error bars represent standard error of the mean. Asterisk (*) indicates significantly more error than HUD.

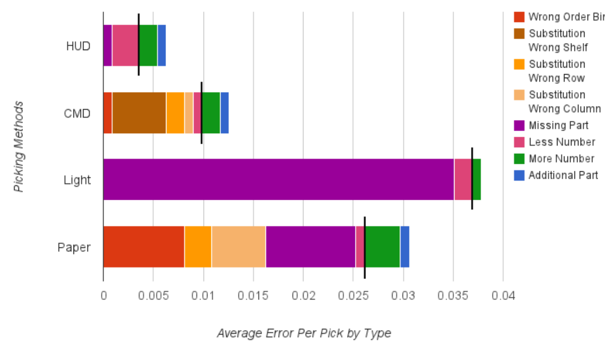


Figure 8: Average error per pick by type.

HUD	A			B			Order Bins		
	1	2	3	1	2	3	1	2	3
1		1	1		1				
2	1								
3			1						
4			1	1			Total: 7		

CMD	A			B			Order Bins		
	1	2	3	1	2	3	1	2	3
1						1		1	
2	1		4		1				
3		1	2						
4	1		1			1	Total: 14		

Light	A			B			Order Bins		
	1	2	3	1	2	3	1	2	3
1	5	1	2	3		11			
2	2		1	2		5			
3	1		1	2		1			
4			2	3			Total: 42		

Paper	A			B			Order Bins		
	1	2	3	1	2	3	1	2	3
1		1			2	2	5	3	1
2		3		1	2	1			
3	1					2			
4	2	2	1	2	2	1	Total: 34		

Figure 9: Positions where errors occurred for all four picking methods.

was used to compare the average task time for each of the picking methods. The average task time using HUD ($M = 38.570, SD = 4.358$) was significantly faster than using light ($M = 45.507, SD = 6.180$), $t(7) = -4.383, p = 0.002$ (one-tailed), and paper ($M = 62.270, SD = 14.100$), $t(7) = -6.684, p = 0.0001$ (one-tailed).

Figure 11 shows the average time required for participants to complete the last eight tasks in the testing session for each of the four picking methods. The task times are generally consistent over these final eight tasks, indicating that participants are no longer increasing their speeds due to learning effects.

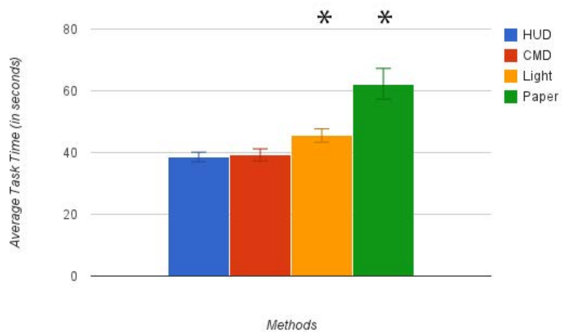


Figure 10: Average task time. The error bars represent the standard error of the mean. An asterisk (*) indicates a significantly slower task time than HUD.

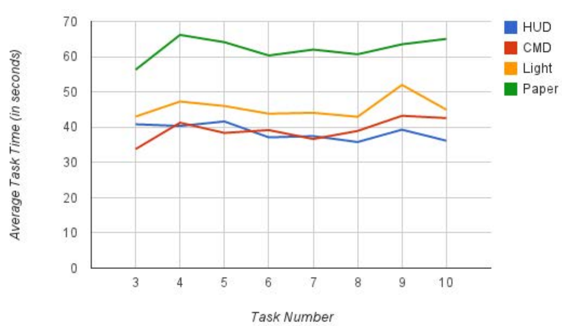


Figure 11: Average task time, by task number.

Overall Task Load

We used NASA-TLX forms to measure the cognitive load of each system. The overall task load for each of the picking methods are presented in Figure 12. A one-tailed paired samples t-test was used to compare the overall task load for each method. The overall task load using HUD ($M = 3.808, SD = 1.931$) was significantly lower than using light ($M = 5.679, SD = 0.870$), $t(7) = -2.730, p = 0.015$ (one-tailed), and paper ($M = 7.008, SD = 0.986$), $t(7) = -3.266, p = 0.007$ (one-tailed).

User Preferences

The median user preference rankings for overall preference, learnability, comfort, speed, and accuracy for all four picking methods are shown in Table 2. The ranks were compared

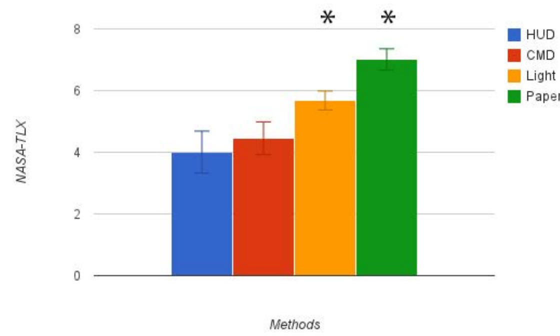


Figure 12: Overall task load (NASA-TLX). Error bars represent the standard error of the mean. An asterisk (*) indicates a task load significantly higher than pick-by-HUD.

using a Wilcoxon Signed Rank Test, the nonparametric equivalent of a paired samples t-test. The HUD ($Md = 1.0$) was ranked significantly higher overall than light ($Md = 3.0$), $z = -2.598, p = 0.005$ (one-tailed), with a large effect size ($r = -0.919$), and paper ($Md = 4.0$), $z = -2.598, p = 0.005$ (one-tailed), with a large effect size ($r = -0.919$).

The HUD ($Md = 2.0$) was ranked significantly higher than paper ($Md = 4.0$) with regards to learnability, $z = -1.919, p = 0.028$ (one-tailed), with a large effect size ($r = -0.678$).

On the speed measure, the HUD ($Md = 1.0$) was ranked significantly higher overall than the other three order picking methods: CMD ($Md = 2.0$), $z = -2.111, p = 0.017$ (one-tailed), with a large effect size ($r = -0.746$), light ($Md = 3.0$), $z = -1.852, p = 0.032$ (one-tailed), with a large effect size ($r = -0.655$), and paper ($Md = 4.0$), $z = -2.226, p = 0.013$ (one-tailed), with a large effect size ($r = -0.787$).

When asked to rank each of the methods in order of resulting accuracy, the participants ranked the HUD ($Md = 2.0$) significantly better than CMD ($Md = 2.5$), $z = -2.121, p = 0.017$ (one-tailed), with a large effect size ($r = -0.750$), and paper ($Md = 4$), $z = -2.126, p = 0.017$ (one-tailed), with a large effect size ($r = -0.752$).

	HUD	CMD	Light	Paper
Overall	1	2	3 *	4 *
Learnability	2	2	2	4 *
Comfort	2.5	1.5	2.5	4 *
Speed	1	2 *	3 *	4 *
Accuracy	2	2.5 *	2.5	4 *

Table 2: User preferences. An asterisk (*) indicates a significantly lower ranking than pick-by-HUD.

DISCUSSION

Of the four picking methods we evaluated, pick-by-paper and pick-by-light are currently used in industry. Pick-by-paper is the most common current practice and is cheap, requires low maintenance cost, and is easy to change. Pick-by-light, while faster than pick-by-paper, is expensive for both initial setup and subsequent reconfiguration.

Our evaluation also included pick-by-HUD and pick-by-CMD, and confirmed our a priori hypotheses that pick-by-HUD offers superior performance to paper and pick-by-light methods on all metrics, including speed, accuracy, error types, workload, and user preferences. Surprisingly, there were few metrics that showed a statistical difference in pick-by-CMD and pick-by-light. Pick-by-CMD scored almost as well as pick-by-HUD on all metrics.

These results suggest that pick-by-HUD and pick-by-CMD are promising competitors for current industrial applications. Both should be more cost-effective than pick-by-light in terms of initial setup costs and flexibility in reconfiguration.

The participants were mostly accurate in evaluating their own performance when using the four picking methods. Participants felt that they were fastest using the HUD, then CMD, then light, then paper, which is consistent with the actual average task time. Participants also felt that HUD was more accurate than the other methods, which is consistent with the actual results. However, the participants felt that light was as accurate as CMD, and was better than paper, when in fact the opposite was true. This conflict between perception and reality is probably because the participants did not even notice the items they missed with pick-by-light, which is a negative indication for the real-world accuracy of this method.

Critiquing Different Systems

With pick-by-paper, users need to use one hand to hold the paper pick list. Thus they normally only have one free hand to pick, greatly reducing the speed of picking. Also, due to the limit of number of items they can hold at once, they may need to split the subtask into multiple parts.

Because the entire pick-by-paper task is presented at one time (including all subtasks), users can optimize by performing multiple subtasks at the same time. This strategy occurs when several nearby subtasks contain few pick items.

The possible reason for so many errors with pick-by-paper is that it requires a heavy memory load, as the results show in Figure 12, and there is not a natural mapping from the paper pick lists to where to pick.

One participant performed pick-by-paper from memory more than any of the other participants. He read several lines on the list first and memorized it, then put the list away and used both hands to do the pick, even performing optimizations while memorizing. Possibly as a result of this memory strategy, this participant did not commit any picking errors.

Pick-by-light is faster than pick-by-paper, in part because the users can employ both hands to perform the picks. However, with pick-by-light users need to visually scan the shelving unit from top left to bottom right which takes time. They also tended to step back frequently to see the entire shelving unit, because they could not keep the complete context in their heads.

Pick-by-CMD is inexpensive and fast. As with pick-by-light, the instrumentation is in the environment rather than on the user, which may increase comfort. However, users may need to turn theirs heads between the cart-mounted display and the order bins. If their focus is mainly on the display, they may need to use peripheral vision to pick and place.

Pick-by-HUD is fast mainly because users can employ both hands, and the display is always available. Users do not need to step back to see the entire shelving unit, as they would for pick-by-light or pick-by-vision (Augmented Reality) [8], and they do not need to turn their heads to alternately look at the display and the bins.

For pick-by-HUD, we observed that users occasionally adjust the HUD or lean and turn their heads toward the HUD side, possibly because they are not accustomed to the head-up display. Also, they appeared to sometimes not move their heads as naturally as normal, possibly because of comfort issues. They may be afraid that the HUD will fall off or move. For participants who did not wear glasses, the HUD appeared easier to wear.

Although the participants in the study did not necessarily have previous experience with HUDs, pick-by-HUD is not harder to learn than any of the other methods according to the participants' stated preferences. In fact, pick-by-HUD was significantly easier to learn than pick-by-paper.

With our current testing system, participants had to wear a backpack with a wearable computer in it, which may not be necessary as HUD technologies improve. However, according to participant preferences, HUD (including the backpack) was no less comfortable than light, and was significantly more comfortable than paper. Previous work has shown that pick-by-HUD in real-world warehouses can be used by workers continuously over weeks without difficulty [3]. Improvements in miniaturization (i.e., Google Glass) may further improve comfort of pick-by-HUD.

Error Types

From Figure 8 and Figure 9, we can see a variety of error types for the four picking methods. Wrong order bin errors happened more frequently for pick-by-paper than for the other picking methods. These errors are likely due to the high task load and memory required by this method. After the participants finished picking from the order bins, they may have forgotten which subtask they were performing.

For pick-by-light, substitution and wrong bin errors never happened, likely because the mapping of LED lights to position is very natural and intuitive. However, most of the errors were missing part errors, which means that participants skipped those pick bins. Also, these errors all happened on the edges of the shelving units, where the participants did not have a clear view unless they stepped back to view the entire shelving unit.

For pick-by-CMD, errors were spread out into a wide variety of error types. Substitution - wrong shelving unit errors happened frequently and only happened in pick-by-CMD. This error mostly happened when the participant finished the final subtask on shelving unit B. Then, when the new task was presented, the participant started to pick from shelving unit B, rather than moving to shelving unit A.

If we compare HUD and CMD regarding error rates and types, we find that no substitutions errors happened on HUD, but many substitutions happened on CMD. This contrast is probably because with HUD, the participants get a better view of the shelf, and do not need to look back and forth as with CMD.

Considering the severity of errors, where more number and additional part errors are less severe, a smaller proportion of errors will stop an assembly line: 4/7 in HUD (0.3% errors per pick) versus 11/14 (1% errors per pick) in CMD.

There were more errors for light than paper (Figure 7). Our system did not include a feedback and error correction system, as is found in many industrial systems. Including one may reduce errors, but will also increase the average task time due to the extra time spent correcting these errors. Real-world pick-by-light systems also typically use brighter LEDs and require a confirmation of each pick. By contrast, our experimental pick-by-light system requires only a confirmation of each place, in order to be consistent with the CMD and HUD pick methods.

Accuracy

Adding weight sensors or scales under order bins would help to reduce the error rate. Such a scale feedback system is much more effective and cheaper than a laser scanner, which is more than 1000 € to upgrade one meter of a pick line [2]. As fixing the errors takes a great deal of time and pick-by-HUD showed the fewest errors, the combination of pick-by-HUD with a scale feedback system would perform the best.

We noticed that when the experimenter forgot to proceed to the next pick, users occasionally made the mistake of doing the same pick again during pick-by-light. In other methods, such as pick-by-CMD or pick-by-HUD, the users would wait for, or even remind, the experimenter. This effect is caused by the lack of transition between picks and tasks. However, as these mistakes are caused by experimenters instead of the users, we did not count them as errors.

In order to better approximate real-world picking scenarios, we attempted to complicate the tasks for participants by making them less routine and predicable for the participants. Presumably, this attempt would have increased the error rates above that observed in previous work. However, our error rates were only slightly higher, in the 0.6% to 3% range, as compared to 0.4% to 1.9% range [10].

CONCLUSION AND FUTURE WORK

Our empirical evaluation shows that pick-by-HUD and pick-by-CMD are superior on all metrics to the current practices of pick-by-paper and pick-by-light, but the differences between HUD and CMD are not significant and do not show that HUD is better than CMD.

One possible reason is that we are not using the best HUD for the task. We will conduct another study comparing CMD and three different HUDs (SV-3, Epson Moverio, and Google Glass). The study will not only provide us with results of different order picking methods, but also show the minor differences of different HUD systems (display position up or down, transparent or opaque, one-eye or two-eye), and the sensitivity of order picking tasks to these various displays. Because the difference between HUD and CMD is small in the study in this paper, it will be useful to recruit more participants for the new study.

Other future work includes an evaluation of error reduction techniques. In one such system, we would add a scale feedback system to pick-by-HUD. We could also compare pick-by-light alone, pick-by-light with a laser scanner, and pick-by-light with a scale feedback system.

REFERENCES

1. Baumann, H. *Order picking supported by mobile computing*. PhD thesis, Doctoral Dissertation. University of Bremen, 2013.
2. Baumann, H., Starner, T., Iben, H., Lewandowski, A., and Zschaler, P. Evaluation of graphical user-interfaces for order picking using head-mounted displays. In *ICMI*, ACM (2011), 377–384.
3. Baumann, H., Starner, T., and Zschaler, P. Studying order picking in an operating automobile manufacturing plant. In *ISWC*, IEEE (2012), 112–113.
4. De Koster, M. Warehouse assessment in a single tour. In *Warehousing in the Global Supply Chain*. Springer, 2012, 457–473.
5. De Koster, R., Le-Duc, T., and Roodbergen, K. J. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research 182*, 2 (2007), 481–501.
6. 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.
7. Iben, H., Baumann, H., Ruthenbeck, C., and Klug, T. Visual based picking supported by context awareness: comparing picking performance using paper-based lists versus lists presented on a head mounted display with contextual support. In *ICMI*, ACM (2009), 281–288.
8. Reif, R., Günthner, W. A., Schwerdtfeger, B., and Klinker, G. Pick-by-vision comes on age: evaluation of an augmented reality supported picking system in a real storage environment. In *Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, ACM (2009), 23–31.
9. Tompkins, J., White, J., Bozer, Y., and Tanchoco, J. Facilities planning, 2003.
10. 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 *CHI*, ACM (2010), 1695–1704.