A Comparison of Order Picking Methods Augmented with Weight Checking Error Detection

Xiaolong Wu[1], **Malcolm Haynes**[1], **Anhong Guo**[2], **Thad Starner**[1]

[1]Georgia Institute of Technology
85 Fifth Street NW
Atlanta, GA 30332
{xwu86, mghaynes, thad}@gatech.edu

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

ABSTRACT

Order picking accounts for over 50% of the annual $150 billion spent in warehouse operations in the United States. 80% of order picking is performed manually. We compare the industry default of unassisted paper-list picking and three techniques augmented with weight checking: pick-by-light, pick-by-HUD (head-up display), and a combination of light and HUD. HUD+light had the fastest picking (19% faster than paper) and a low number of errors (67% fewer than paper).

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 process of retrieving items from a storage location (*e.g.*, a warehouse) to fulfill customer orders. Order picking is expensive. A recent report estimated that annual warehousing costs in the United States are $148 billion [9], with order picking accounting for between 50%-75% of that expense [3]. A small improvement in order picking efficiency can have a large economic impact.

Despite advances in automation, order picking is often a manual task. Research in the wearable and ubiquitous computing community has focused mainly on improving pick methods via use of head-mounted displays and augmented reality [4, 5, 6, 7, 8]. These methods vie with established methods such as pick-by-light and pick-by-voice to maximize pick speed and accuracy. However, while the aforementioned methods all attempt to reduce pick errors, few are able to detect them.

Weight checking is an error detection method that can augment pick methods by determining the quantity of an item type based on weight. Weight checking (combined with bar code scanning) is used in many self-checkout lanes at retail stores. However, despite the prevalence of weight checking, little research examines the efficacy of weight checking as a technique to augment traditional order picking methods.

We compare four pick methods: pick-by-light, pick-by-HUD (head-up display), and pick-by-HUD with light, all augmented with weight checking, and the industry default of an unaugmented paper pick list. The paper pick list method acts as a reference task across most of our previous experiments [5, 6]. The pick-by-HUD with light method introduced here optimizes both speed and accuracy, showing that businesses already invested in a light system can further improve performance by integrating a HUD system with weight checking.

RELATED WORK

Studies on order picking methods often analyze the speed and accuracy of different methods. Guo et al. compared pick-by-paper, pick-by-light, pick-by-display, and pick-by-HUD and found that pick-by-HUD was faster than the other pick methods [5, 6]. Reif et al. developed pick-by-vision, a pick method that uses augmented reality, and compared it with pick-by-paper [8]. Ali et al. used a scale under each item bin to determine if a user picked the correct quantity of an item [1]. Baechler et al. used a single scale implemented at the order receiving cart to determine if the user picked the correct quantity of an item type [2]. Companies such as Bruss and Avery Weigh-Tronix have developed order picking systems that incorporate weight checking for industry applications.

PICK METHODS

All pick implementations are consistent with the study procedures described by Wu et al. [10], using Google Glass for the HUD and a pick-by-light system constructed to imitate ones observed in automobile manufacturing environments. We augmented all pick methods, except pick-by-paper, with our weight checking system. We tested all systems using a dense pick environment designed to induce pick errors.

Pick-by-HUD with light

In addition to the use of weight scales, we introduce pick-by-HUD with light. In previous studies, pick-by-HUD was significantly faster than pick-by-light, but there was a trend toward more errors (that was not significant statistically). We hypothesize that errors may stem from subjects memorizing portions of the order so they could shift attention from the HUD to the item bins while picking. Because short-term memory is limited, users tended to make mistakes, especially on orders where four or more bins were picked per shelving unit. We hypothesize adding a real-world visual indicator of

ISWC '16, September 12-16, 2016, Heidelberg, Germany
© 2016 ACM. ISBN 978-1-4503-4460-9/16/09...$15.00
DOI: http://dx.doi.org/10.1145/2971763.2971773

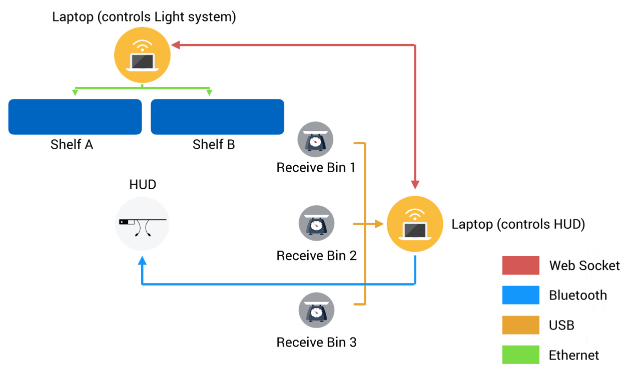


Figure 1: System architecture.

the correct bin and quantity (via the light system) will improve errors while maintaining speed. In pick-by-HUD with light, the picker wears a HUD that works the same way as in the pick-by-HUD method. In addition, the picker is given another visual aid in the form of lit LEDs that work in the same way as the pick-by-light method. However, button push pick confirmation is not required.

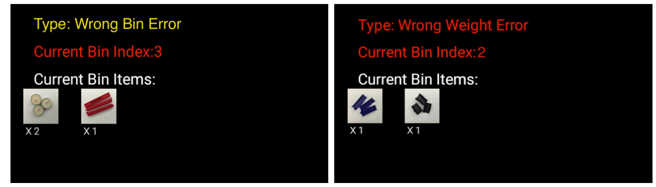
IMPLEMENTATION

We implement weight checking using scales which provide input to an error detection algorithm. If the algorithm determines the weight is correct, the system automatically advances to the next order. Otherwise, the system displays an error message if the weight is incorrect. An overview of the system architecture is shown in Figure 1.

For load sensors, we use Dymo 10-pound digital postal scales with weight sensitivity of 0.1 oz. (2.8 g). The load sensors connect via USB to an Acer Chromebook computer running Ubuntu Linux 14.04. Each of the three order receive bins used in our pick environment is placed on a scale. The scales are then calibrated to zero so that the weight of the empty order bin is removed from the weight reported by the scales.

The error detection algorithm calculates the expected value for the weight of the current order based on the known weight of items in the order. If the bin weight falls within an acceptable range, the algorithm accepts the pick as correct. Otherwise, the algorithm rejects the pick as incorrect. The system deems a pick incorrect for two reasons: the user either placed items in the wrong bin (wrong bin error) or placed the wrong items in the right bin (wrong weight error).

If the system detects an error, it displays an error message. The HUD system graphically displays either error type - wrong bin or wrong weight - as shown in Figure 2. The light system displays wrong weight errors by resetting the LEDs associated with each item type in the order to the quantity of the item type that should be picked. It resets the LED associated with the receive bin to the total number of items that should be placed in the receive bin. The system then repeatedly flashes these LED numbers approximately three times per second. For wrong bin errors, the light system simply does not advance to the next order.



(a) Wrong bin error.

(b) Wrong weight error.

Figure 2: Error message on HUD.

Users can dismiss errors in two ways. First is to fix the error. When the correct items are placed in the bin, the system detects the correct weight, dismisses the error, and advances to the next task. The user can also ignore the error and manually advance to the next task by pushing the receive bin button.

STUDY METHODOLOGY

We conducted a within-subjects user study to evaluate the pick methods. Our study consisted of 12 participants ages 20-33 (M = 24.2) from our university (7 male, 5 female). Nine participants were right-eye dominant, and three were left-eye dominant. All participants were first-time order pickers but were given sufficient practice tasks to extinguish learning effects. The pick methods were counterbalanced using a Latin square to avoid ordering effects. Participants were paid $20 for their participation in the study.

Participants were told to complete orders as quickly and accurately as possible. Picked items for an order were placed in the order bin all at once. During a practice session, researchers manually induced errors so that participants experienced the error detection and correction system. Test sessions were video recorded and timed. Participants took a NASA TLX survey for each pick method and a preference survey at conclusion of testing.

ANALYSIS AND RESULTS

We report average task time, average error per pick, error types, subjective task load, and user preferences for all four pick methods. Based on previous research, we held the following hypotheses:

- Pick-by-HUD and pick-by-HUD with light will be significantly faster than the other pick methods;
- Pick-by-HUD and pick-by-HUD with light would generate lower workload than pick-by-paper;
- Pick-by-HUD with light would be ranked significantly higher than the other pick methods for user preferences.
- Pick-by-HUD with light will have less errors than pick-by-HUD, and all methods would have less errors than paper.

Average Task Time

The average task time for each of the pick methods is presented in Figure 3. A one-way ANOVA with repeated measures with Bonferroni correction was used to compare the average task time for each of the pick methods.

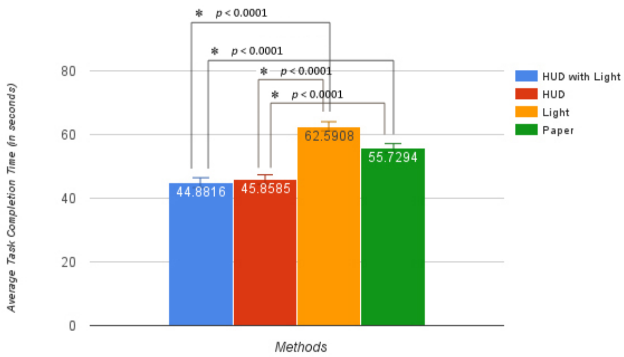


Figure 3: Average task time. *=statistically significant.

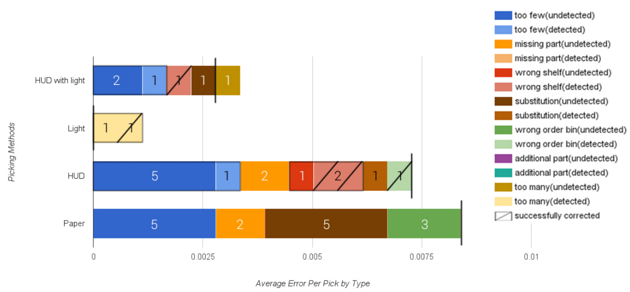


Figure 4: Average error per pick by type.

As hypothesized, the average task time using pick-by-HUD with light ($M = 44.9$ sec, $SD = 13.6$) and pick-by-HUD ($M = 45.9$ sec, $SD = 13.6$) were significantly faster than pick-by-paper ($M = 55.7$ sec, $SD = 14.1$) and pick-by-light ($M = 62.6$ sec, $SD = 14.8$)($p < 0.0001$). In post-hoc analysis, pick-by-paper was significantly faster than pick-by-light ($p = 0.0005$), which was expected given previous research.

Errors
We used a one-way ANOVA with repeated measures with Bonferroni correction to analyze the error data. The error results are shown in Figure 4. There is no statistical difference among the four pick methods; as with previous studies, the number of errors was too small for clear conclusions to be made. Pick-by-paper generated the highest error rate, with an average of 0.84% error per pick (15 total); pick-by-HUD resulted in an average of 0.56% error per pick (10 total); pick-by-HUD with light resulted in an average of 0.28% error per pick (5 total); and pick-by-light resulted in the lowest error rate, with an average of 0.06% error per pick (1 total).

Our error detection system detected 2 errors for pick-by-light, and participants fixed 1 successfully. The system found 5 errors for pick-by-HUD; participants succeeded in correcting 3. For pick-by-HUD with light, the system detected 2 errors, and participants succeeded in correcting 1. For error types, our weight checking system detected 100% of wrong order bin errors. In addition, it detected when the user picked a wrong item 66.7% of the time, too many items 66.7% of the time, and too few items 22.2% of the time.

Average Task Workload
The average task workload was analyzed by a one-way ANOVA with repeated measures with Bonferroni correction. As hypothesized, pick-by-paper ($M = 61.5$, $SD = 21.7$) generated significantly higher workload than pick-by-HUD ($M = 39.4$, $SD = 18.0$, $p = 0.01$), and pick-by-HUD with Light ($M = 39.4$, $SD = 18.0$, $p = 0.04$).

User Preferences
We used the Wilcoxon Signed Rank Test to analyze participants' preferences. We examined their overall preference, learnability, comfort, perceived speed, and accuracy. Our hypothesis was that pick-by-HUD with light would be preferred in all categories; other comparisons reported below are post-hoc.

As hypothesized, for overall preference, pick-by-HUD with light ($Md = 1.0$) was ranked significantly higher than pick-by-HUD ($Md = 2.0$, $p = 0.01$), pick-by-light ($Md = 3.0$, $p = 0.0015$), and pick-by-paper ($Md = 4.0$, $p = 0.0005$).

With regards to learnability, there is no statistically significant difference found among the four approaches.

For comfort, pick-by-paper ($Md = 4.0$) was ranked significantly lower than pick-by-light ($Md = 1.0$, $p = 0.001$), pick-by-HUD ($Md = 2.0$, $p = 0.0045$), and pick-by-HUD with light ($Md = 2.0$, $p = 0.0051$).

For perceived speed, pick-by-HUD with light ($Md = 1.0$) was ranked significantly higher than pick-by-light ($Md = 3.0$, $p = 0.0015$), and pick-by-paper ($Md = 4.0$, $p = 0.001$). There was no statistical difference between pick-by-HUD with light and pick-by-HUD ($Md = 2.0$).

In terms of user perceived accuracy, pick-by-HUD with light ($Md = 1.0$) was ranked significantly higher than pick-by-HUD ($Md = 2.0$, $p = 0.0025$), pick-by-light ($Md = 3.0$, $p = 0.0014$), and pick-by-paper ($Md = 4.0$, $p = 0.0005$). There was no statistical difference between pick-by-HUD and pick-by-light.

DISCUSSION AND FUTURE WORK
Our results are broadly consistent with results from previous studies for pick speed. However, the relative number of errors made with HUD compared to the paper reference method were more than our previous studies. Half of the errors (6 out of 13 from the HUD condition) were made by two participants who had difficulty seeing the screen. Even so, as hoped, weight checking was successful in detecting errors.

Weight Checking Error Detection Limitations
The obvious limitation with our weight checking error detection system is that it cannot detect differences between items with no or negligible weight differences. For example, many real world items come in a variety of colors. These items will weigh exactly the same amount.

Another limitation of our weight checking error detection system stems from variations in the item manufacturing process. Items that are ostensibly the same may weigh slightly

	Light	HUD	HUD with light
No Error / Error Ignored	61.2	43.6	43.2
Error Corrected	125.8	83.4	117.6

Table 1: Average completion times (in seconds) with and without error detection.

different amounts. Hence, the weight detection algorithm accepted as correct order weights that varied within a range. As a result, missing or extra items that collectively keep the bin weight within the acceptable weight range will not trigger an error. This phenomena accounts for every error not detected by our system.

HUD Improvement and Error Recovery

Even though our system identified errors, we found that users sometimes intentionally ignored the error and continued with the pick. One reason users ignored errors was that they were difficult to fix. In general, it took users more than twice as long to complete a task with an error than to complete a task without an error (Table 1).

A major reason users had issues recovering from errors is because the weight checking system was not able to tell the user what specific items were missing from, or extraneous to, the pick. Therefore, to fix the error, the user had to look in the order bin and compare what was in the bin with what should be in the bin. For large orders this process was tedious, and users felt pressure to complete the task quickly. During exit interviews, users indicated that it was difficult to examine every item in the bin to determine what was missing even when the system told them they made an error. Therefore, future work in this area is to improve the detection system so that it suggests likely missing or excess items to the user.

Speed of Pick-by-Light

When cost is not an issue, pick-by-light is a preferred pick method in industry. However, our pick-by-light implementation was slower than pick-by-paper. Users had difficulty fixing detected errors. The additional time spent fixing an error increases average task time. Also, our dense pick environment slowed users as they often had to carefully scan both vertically and horizontally to find the next lit item. In some industrial settings, items to be picked tend to be more sparsely distributed and spread horizontally, making it easier to identify the next item to pick.

HUD, Light, and Visual Attention

Users applied one of two distinct strategies for HUD picking. In one strategy, users shifted visual attention between the HUD, to learn what items to pick, and the item bins, to locate and pick items. In the second strategy, users took advantage of our dense pick environment and only focused on the HUD, picking from the bins by feel.

Adding light to HUD helped participants who shifted attention between the HUD and item bins. Perhaps the light removed the need to memorize item quantities, thus reducing errors. Users who focused solely on the HUD did not benefit from light. As an example, after completing the HUD with light session, one user asked if the item bin lights were ever turned on. Perhaps instructing pickers to use the shifting focus strategy might have reduced errors further.

CONCLUSION

We augmented previously studied order picking methods with weight checking. Weight checking did detect errors; however, recovery from errors was time consuming, and users often abandoned efforts to fix errors due to the difficulty of the task.

Of the pick methods examined, the pick-by-HUD methods provide a good combination of speed and accuracy. In addition, they are preferred over light and paper and produce a lower workload on users.

REFERENCES

1. Ali, S., Lewandowski, A., and Rett, J. A SOA based context-aware order picking system for warehouses using laser range finder and wearable computer. In *WoWMoM*, IEEE (2011), 1–8.
2. Baechler, A., Baechler, L., Autenrieth, S., Kurtz, P., Hoerz, T., Heidenreich, T., and Kruell, G. A comparative study of an assistance system for manual order picking–called pick-by-projection–with the guiding systems pick-by-paper, pick-by-light and pick-by-display. In *HICSS*, IEEE (2016), 523–531.
3. Coyle, J. J., Bardi, E. J., Langley, C. J., et al. *The management of business logistics*, vol. 6. West Publishing Company St Paul, MN, 1996.
4. Funk, M., Shirazi, A. S., Mayer, S., Lischke, L., and Schmidt, A. Pick from here!: an interactive mobile cart using in-situ projection for order picking. In *IJCPUC*, ACM (2015), 601–609.
5. Guo, A., Raghu, S., Xie, X., Ismail, S., Luo, X., Simoneau, J., Gilliland, S., Baumann, H., Southern, C., and Starner, T. A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list. In *ISWC*, ACM (2014), 71–78.
6. Guo, A., Wu, X., Shen, Z., Starner, T., Baumann, H., and Gilliland, S. Order picking with head-up displays. *Computer 48*, 6 (2015), 16–24.
7. Li, X., Chen, I., Thomas, S., and MacDonald, B. Using kinect for monitoring warehouse order picking operations. In *Aust. Conf. on Robotics and Auto.* (2012).
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 *AFIGRAPH*, ACM (New York, NY, USA, 2009), 23–31.
9. Wilson, R. State of Logistics Report. *The Council of SCM Professionals* (2015).
10. Wu, X., Haynes, M., Zhang, Y., Jiang, Z., Shen, Z., Guo, A., Starner, T., and Gilliland, S. Comparing order picking assisted by head-up display versus pick-by-light with explicit pick confirmation. In *ISWC*, ACM (New York, NY, USA, 2015), 133–136.