Visual Based Picking Supported by Context Awareness

Comparing Picking Performance Using Paper-based Lists Versus List Presented on a Head Mounted Display with Contextual Support

Hendrik Iben,
Hannes Baumann
Center for Computing Technologies (TZI)
University of Bremen
Am Fallturm 2
D-28359 Bremen, Germany
{hiben, hannes}@tzi.de

Carmen Ruthenbeck
BIBA - Bremen Institut für Produktion und Logistik GmbH
Hochschulring 20
D-28359 Bremen, Germany
rut@biba.uni-bremen.de

Tobias Klug
SAP AG
Bleichstrasse 8
64283 Darmstadt, Germany
t.klug@sap.com

ABSTRACT

Warehouse picking is a traditional part of assembly and inventory control, and several commercial wearable computers address this market. However, head mounted displays (HMDs) are not yet used in these company's products. We present a 16 person user study that compares the efficiency and perceived workload of paper picking lists versus a HMD system aided by contextual cueing. With practice, users of the HMD system made significantly faster picks and made less mistakes related to missing or additional picked items overall.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Representation**]: User Interfaces—*benchmarking, theory and methods, screen design*

General Terms

wearable computing, benchmark methodology, user study

Keywords

picking, hmd

1. INTRODUCTION AND RELATED WORK

Early in the history of wearable computing, researchers focused on the tasks of inspection, maintenance, and repair as potential areas where wearable computing might prove beneficial [9, 7]. Siewiorek *et al.* provide an overview of the lessons learned from user studies of deployed prototypes in these areas [8]. Commercially, one early success by Symbol Technologies (acquired by Motorola) was in creating a arm-mounted barcode scanner that could speed package scanning and inventory control [11]. Pittsburgh-based Vocollect addressed another niche, inventory picking, using their speech-only interface [10]. While Vocollect has been successful, the user studies mentioned above for similar problems suggest that head-mounted displays (HMD) might also prove useful for the task of inventory picking. Our work here focuses on the inventory picking problems faced by our partners in the automobile industry.

Picking is the process of collecting items from an assortment in inventory and represents one of the main activities performed in warehouses. Picking accounts for 55% [1] to 65% [3] of the total operational costs of a warehouse. Typically the process begins with a picking list, which 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. A worker collects the items from stock and transports the items to a specific location.

In automobile manufacture, our specific focus, engines are assembled from various inventory stock based on model number. Errors in picking can slow the assembly process considerably. Meanwhile, manufacturers want the picking process itself to be as efficient as possible. Yet, the various sizes and shapes of the parts make automation with robotics difficult.

Like Mizell's augmented reality task of assembling wire bundles for aircraft [5], inventory picking is well-suited for experimentation. Most laboratories that make physical prototypes experience inventory problems related to those in the production environment (Yarin and Ishii [12] take advantage of this fact and demonstrate the use of contextual information in a picking scenario by providing ambient information on shelf usage). Thus, inventory picking environments are easily reproduced in the laboratory (with varying levels of fidelity) where conditions can be controlled and the user can be monitored.

Besides being readily reproduceable and modeled after a real-world task, experimental procedures to explore wearable computer designs should allow ease of variation of conditions, such as type of subject, equipment, and interface. Small changes can often have large effects, and the researcher needs to uncover what factors are most important in user performance.

Ideally, we would like an experimental environment where

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.

a wearable computer interface can demonstrate improvements over traditional methods, and the metrics used to evaluate the system have sufficient sensitivity to show the effect of changes to the interface. Given past commercial successes, picking seems a logical choice. Picking has the benefits of being simple to teach to novice users and fast to perform so that many trials may be executed in a small period of time. Hopefully, learning effects can be modeled, and expert users trained quickly, if desired. In addition, picking can be made mentally, visually, and manually taxing, similar to many other tasks that are being investigated for augmentation with wearable computers. Standard quantitative metrics, such as performance time and accuracy, and subjective metrics, like the NASA Task Load Index (TLX) [4], can be applied to experiments. Finally, picking also has a simple, ecologically-valid control condition that can be used for comparision — the paper picking list.

In this paper, we share our attempt at creating an inventory picking experimental procedure, describe potential improvements, and provide the results of a study that compares the use of a traditional paper-based picking list to a wearable-based system with HMD and contextually-based cueing [2].

2. EXPERIMENTAL ENVIRONMENT

We designed the experimental environment and the task to closely model that of the current assembly processes of our partners. For convenience, we will use the term "pick" to mean the removal of one or more items from an inventory bin. Subjects might pick several items before placing them into the specified receiving bin on the picking cart (see figure 2). We group picks into tasks. For each task, the user must pick a variety of parts from the inventory bins and place them in the correct receiving bins. As many as eight receiving bins may be used representing up to eight "pre-assembly" groups that the subject must pick. In keeping with our observation of actual work situations, some pre-assembly groups require a small amount of manual assembly before the task is considered complete. In practice, such assemblies provide real-time plausibility check to the user as to the correctness of his picks. For example, some of the parts assembled into portions of a small house. Figure 3 shows the result of several of these manual assemblies. For convenience and reproducibility, we use the popular LEGO^{TM} brand building blocks for our parts.

2.1 Inventory and receiving bins

Figure 1 shows an overhead view of the experimental setting. The user stands with his picking cart in four shelves, labeled A-D. Each shelf consists of 12 bins. Each bin is labeled with its row and column position on the shelf. The subject places picked items into receiving bins on a wheeled cart (see figure 2). The subject is free to move and orient the cart for his convenience. The experimenter records the activities on the participant on a TabletPC, which controls the context displayed during the HMD trials.

2.2 Paper picking list

Figure 4 shows an example paper picking list. Parts to be picked are listed in groups that reflect how they will be assembled. This organization is used in the field and aids pickers' intuition as to which part may be needed next. The picker is free to optimize his performance as he sees fit.

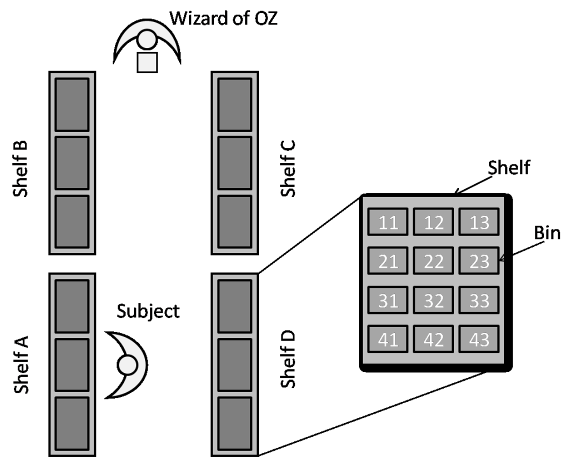

Figure 1: setting sketch

Figure 2: picking cart

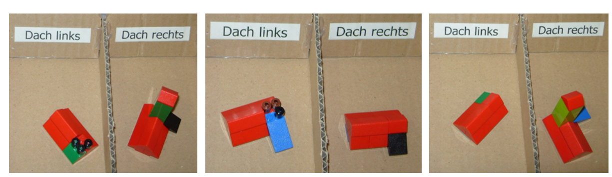

Figure 3: 6 of 74 possible assemblies from 40 tasks

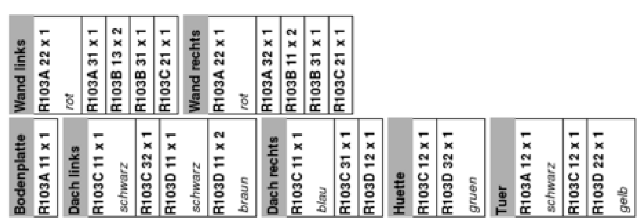

Figure 4: paper pick list for task 9

Figure 5: HMD interface shows which parts need to be picked

Figure 6: the pink highlighting shows the last pick completed

2.3 Wearable computer

For this study, we are using a modified Porteo 60-CX wearable computer made by teXXmo featuring a 1.5GHz x86 processor. The computer is worn in a jacket with the computer stored in a pocket at the back see figure 2). The Trivisio M3 color HMD has 800x600 resolution and is worn over the right eye. The computer is connected by a wireless local area network to the experimenter's tablet PC (a Kaleo GX), and the subject's interface is controlled remotely by the experimenter.

Figures 5, 6 and 8 demonstrate the interface displayed on the HMD. The interface was designed to take advantage of the user's context as they perform the picking task. When the user stands in front of a shelf, only parts that need to be picked from that particular shelf are displayed. The user is presented with the groups of objects to pick. Each part in each group is labeled with its part number, its position on the shelf, the number of parts to be picked, and the color or type to be picked (if applicable). Each group is labeled with the name of the receiving bin (e.g., Bodenplatte, Tuer,

Figure 7: as the user reaches into bin A22 the wizard pressed the appropriate button

Figure 8: A22 is marked completed

Figure 9: error message indicating a wrong pick

Wand links, etc.) into which the parts should be placed. Completed groups are shown in grey and have a green checkmark next to them. Picked parts are also indicated with a green checkmark. Groups to be completed are shown in yellow. The most recent part picked is highlighted in pink so that the user can keep track of his most recent action. If the user reaches into the wrong bin an error message is displayed (see figure 9). The error message is removed when the user reaches into the bin to replace the incorrectly picked part. At the bottom of the display, the name of the current shelf is highlighted. When all parts from a given shelf have been picked, the shelf itself is ticked as complete. When all parts from the task have been picked, a message indication the completion of the task is displayed.

2.4 Wizard of Oz context sensing

As figure 7 shows, the researcher emulates the context sensing system. We choose this "Wizard of Oz" approach for several reasons. First, we wished to see the value of using context before designing and deploying an appropriate sensor system. In addition, we expected to discover new possible uses of context after we gained experience in running the experiment. Furthermore, there are many ways to sense which shelf the user has approached and which bin he is manipulating. Possibilities include RFID tags, accelerometers, proximity sensors, capacitive sensors, and computer vision, among many others. Running a user study before selecting the specific sensors allows us to better determine our requirements for the final system design. Another key advantage is that a Wizard of Oz system is easily reproducible by other researchers who wish to experiment with different hardware or tasks without the overhead of implementing our specific sensing solution. A less obvious advantage is that experimenters can follow the same Wizard of Oz procedures for the experimental condition and the paper-based control condition to capture users' performances. Such is the case in the experiment described below. As the wizard's actions are logged, timing information can be recovered for both the experimental and control conditions afterwards. Since the wizard is performing the same actions for each condition, there is less potential for unforeseen variations in the experiment.

During our experiment, the wizard uses his touchscreen to select which shelf the user has approached. When the user's hand crosses the lip of a shelf, the wizard selects the appropriate bin on his screen. The subject's interface then shows the pick associated with that bin as complete. We do not assume that cost-effective sensors could recognize which part was picked, how many parts where picked, or if the user simply put his hand into the bin and withdrew it without picking a part. Furthermore, we assume the sensors could not determine if a part was replaced. Thus, the wizard's job was fairly simple if strict: select the appropriate shelf and the appropriate bin each time the user's hand crossed the plane of the front shelf. The wizard had to follow these rules very strictly so as not to provide uneven experiences for the different subjects. In addition, in this experiment, we used four wizards, further increasing the importance of carefully following the rules. Note, however, that this simulation results in potentially confusing, but realistic, situations for the users, which we document here. Potential situations where our "context sensors" produce incorrect user feedback include

- **"Sensor" error:** The wizard fails to press a button, presses the wrong button, presses additional buttons, or simply presses a button at the wrong time. If the sensor error caused an error message on the screen, the user might ignore the message or pretend to place an "incorrect" piece in the appropriate bin to remove the message. The user can address the other errors with similar methods.

- **Too few items picked:** For example, the participant picks one instead of two objects for a particular assembly group. This situation is a true error on the part of the subject and is noted at the end of the task. However, no error message is generated by the interface.

- **User picks parts for two assembly groups at once:** Suppose the user sees that the same part number is used in two different assemblies. He might reach into the bin only once and remove enough pieces for all assemblies. The wizard marks this as one reach. Thus, the user's display will not mark as complete the pick for the second assembly. The user might ignore the screen and continue or reach in the bin a second time to force the interface to mark the second pick as complete.

- **User picks the same bin multiple times for the same assembly:** Suppose the user's HMD shows that 2 parts need to be picked for assembly group "Wand links" and 3 of the same parts need to be picked for assembly group "Wand rechts." The user reaches into the bin and picks one part, which results in the pick for "Wand links" being marked as complete. Realizing that he needed two parts for "Wand links", the user reaches into the bin a second time. At this point, the "Wand rechts" pick is incorrectly marked as complete, even though the user has only completed Wand links. The user can not "take back" the pick to correct the system. Instead, he has to remember to complete the remaining pick (providing he realized the problem). This situation can only occur when two or more assembly groups require the same parts, and the most likely result of the situation is that the user forgets the pick for the later assembly group.

3. PILOT STUDY

To avoid unnecessary usability problems with the HMD-system, a pilot study was performed. A total of 15 participants used two variants of the HMD-system and the paper version (5 groups of 3 subjects). The test took about one hour for each participant. We discovered that participants required significant time to become accustomed to wearing the HMD. To combat learning effects, we decided to increase the total time a user spends with the HMD to 2 hours. However, given the resulting long study duration, we changed to a between-subjects study design. We also improved the HMD user interface to address usability problems identified during the study. For example, we changed the colors used in the interface. Due to contrast issues, the original colors were not well suited for a see-through HMD.

4. USER STUDY

25 participants (7 female) were recruited from students of the university with a majority of participants from the department of industrial engineering and computer science. Of these, six were rejected due to early changes in experimental procedure, one was rejected as a non-native German speaker, and one was rejected due to a failure in the timing procedure. Of the remaining 16 subjects, 3 were female. Due to the length of the study, a between-subject design was chosen where the independent variable is the system used (paper or wearable computer). Eight subjects completed each condition. When using the HMD, all subjects were required to wear it over the right eye. Subjects were given an allowance for their time in accordance with the pay scales of student assistants.

Participants were given a sequence of tasks to pick parts from the shelves. 43 different tasks were defined. Three were used for training. The 40 tasks consisted of an average of 22,725 picks. Tasks also included some small assembly problems as described earlier. On average, the 13 tasks in round one required 1.92 assemblies per task, round two required 1.96 assemblies per task, and the average number of assemblies across all 40 tasks was 1.95 assemblies per task. To avoid learning effects regarding the order of tasks, each participant was assigned a different random permutation of the task sequence with the exception that the same order exists once for the HMD and the paper condition.

Paper-based tasks were ordered by pre-assembly groups, as is currently done in industry. Then the list was ordered by shelves and bins. An example paper-based task is shown in figure 4. To take advantage of context awareness, HMD tasks are grouped first by shelf, then assembly, and then bin.

First, each participant performs three tasks for training. These tasks are used to explain the picking scenario and, in the case of the HMD, to adjust it so that the participant can see clearly. Two assemblies needed in the tasks are explained to the participants explicitly. After this learning period, subjects perform two rounds of picking tasks.

The first round consists of 13 picking tasks. The time required for each task is recorded. A photograph is taken of the picked parts in their receiving bins for later error analysis (see figure 10). As the available parts are limited, the picked items are decanted into a box and returned to their inventory bins by an assistant. "Context" information is recorded by the wizard for both the HMD and paper tasks.

The second round is identical to the first with the exception that the subject is given 27 picking tasks. Very few participants finish all tasks before the end of the 2.5 hour period. After the second round, the subjects complete the NASA TLX. Due to a procedural mistake, 11 of the 16 subjects completed the second page ("weightings") section of the TLX, though all subjects completed the scales.

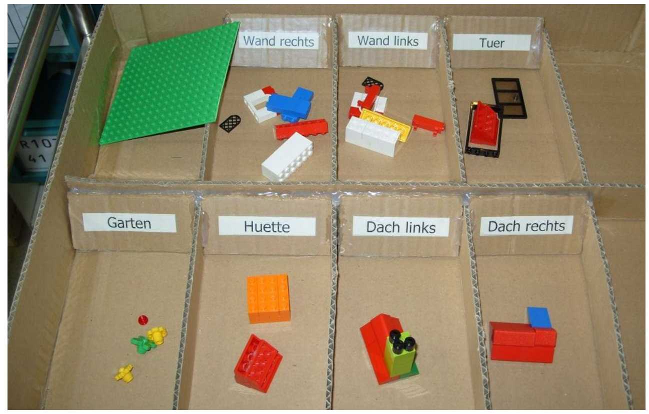

Figure 10: photograph of receiving bins for error analysis

5. RESULTS

In total, subjects completed 524 tasks representing 11,744 correct picks (not including the training tasks). As none of the subjects had used a HMD before the study, much effort in the training tasks and in the first round of tasks related to the subjects learning to use the display properly. However, by the second round of tasks, the HMD users were well accustomed to the display. Of the 27 tasks in the second round, 2 HMD users and no paper users completed all tasks. All subjects completed at least 13 tasks in the second round before their 2.5 hour time slot was completed. When comparing accuracies, we focus on the first 13 tasks in the second round as they represent the more expert use we would expect in practice.

5.1 Speed and Accuracy

We allowed subjects to continue performing tasks in round two until the end of their 2.5 hour time slot. Subjects that used the paper picking list averaged 396 picks in round 2 (ranging from 266 to 494 picks). Subjects using the HMD-system averaged 458 picks (ranging from 337 to 600 picks). While not statistically significant, these numbers are impressive given that, on average, the HMD users spent much more of their 2.5 hour slots learning how to use the system.

On average, subjects using the paper picking lists made 1.1% mistakes per pick (0-8 mistakes total per subject) while users of the HMD system made 0.74% mistakes per pick (0-11 mistakes total per subject). Figure 11 shows these results. We further divide these mistakes into two classes. The first class are mistakes that the context sensitivity of the HMD system could help prevent. These "context mistakes" include picking the wrong item or failing to pick an item. While also a context mistake, no subject took a part from the wrong shelf in the second round (paper users made two such mistakes in the first round).

For context mistakes, paper users made 0.76% mistakes/pick and HMD users 0.19% mistakes/pick. A Welch Two Sample t-test shows that this result is significant ($p < 0.05$). Figure 12 charts the percent of context mistakes made per pick for each subject. The second class of mistakes, which we will name “other mistakes,” include placing a part in the wrong receiving bin, picking the wrong number of parts, and picking the wrong color or type of part. For “other” mistakes, paper users made 0.34% mistakes/pick and HMD users 0.55% mistakes/pick.

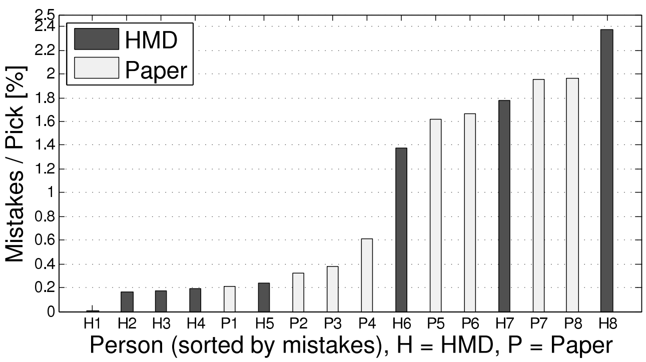

Figure 11: total mistakes made by the subjects.

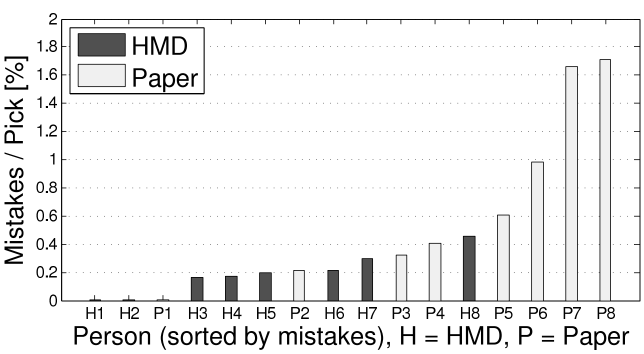

Figure 12: context mistakes: picking the wrong item or failing to pick an item (visible by contextual cues on the HMD).

Given the wizard’s logs, we could calculate the average amount of time required per pick. The time per pick was determined by comparing the time of the last pick (hand movement across a shelf) to the current pick. Picks longer than 25 seconds were ignored as these often correlate to technical problems. Averaging the picks from the 13 tasks from round one and the first 13 tasks from round two, paper users required 6.1 seconds/pick while HMD users required 5.3 seconds/pick. One major advantage of the HMD users was that their picking lists were ordered primarily by shelf. Paper users, concentrating on completing assembly groups that may be across several shelves, may require more time for reorienting themselves to each shelf. If we compare only picks made on the same shelf, paper users needed 4.6 seconds/pick while HMD users required 4.9 seconds/pick. This comparison may reflect an implicit advantage for the paper-based users when picking from only one shelf of parts. Because the paper users picked according to assembly groups, they could gather all parts for an assembly in their hands and then place them in the same receiving bin simultaneously. If HMD users gathered many parts in their hands before transferring them to the cart, they would need to sort the pieces into the correct receiving bin.

A small cause for convern is that one of the paper users picked much faster than any other subject. For picks on the same shelf, subject 10 picked an average of 33.8% faster than other paper subjects. Without this subject’s data, paper users averaged 4.9 seconds/pick for picks on the same shelf – the same as the HMD users. We are investigating this user’s data further as of this writing.

5.2 Learning effects

In figure 13 and 14 the time between picks for for HMD and paper-based users is shown where figure 13 displays times for all picks while figure 14 contains times for picks on the same shelf only.

Figure 13 shows the average time per task over the 13 tasks in round one and the first 13 tasks in round two (for a total of 26 tasks). A learning effect is clearly visible for both paper and HMD users. Fitting a power law curve to the data resulted in good matches. Paper times followed the curve $-15.072x^{0.043} + 22.811$ with $r^2 = 0.852$, HMD times followed $-1.257x^{0.311} + 8.017$ with $r^2 = 0.955$. Looking at the times seen for the same shelf fitting result in the curve $-1.916x^{0.225} + 7.857$, $r^2 = 0.614$ for paper and $-0.949x^{0.346} + 7.082$, $r^2 = 0.927$ for the HMD condition.

While continued learning is evident past task 26, the improvement seems to be approaching an asymptote. The last five tasks (22-26) show a statistically significant difference between the paper and HMD results ($p < 0.05$). The difference over the last five sessions averaged 1.0 seconds/pick. The last three tasks for round one (11-13) also show a statistically significant difference (as well as task 3, 4, and 15). Interestingly, when only picks on the same shelf are considered for the tasks, the results are much more haggard. Figure 14 shows these results.

5.3 NASA TLX

Figure 15 shows the results of the NASA TLX evaluation. The Task Load Index can be used to compare the perceived quality of user interfaces. No statistically significant effect could be found. However, anecdotally, we observed a higher mental demand for many of the HMD users. Figure 15 shows the TLX results with standard deviations.

6. DISCUSSION

6.1 Improving the experimental procedure

Overall, the experimental procedure worked well (after initial tuning) and inventory picking seems a promising reference task for wearable computing researchers. In 2.5 hours, the average subject would perform 734 picks. As seen in figure 13, most learning effects were ameliorated by the end of the study. Effects of the different conditions could be seen quickly and proved significant by the end of each round. However, while the number of errors showed significance where expected, the absolute number of errors were small. Perhaps the task could be made more difficult to better differentiate the conditions.

Many of our particular choices for this experiment were made to preserve ecological validity with the tasks performed

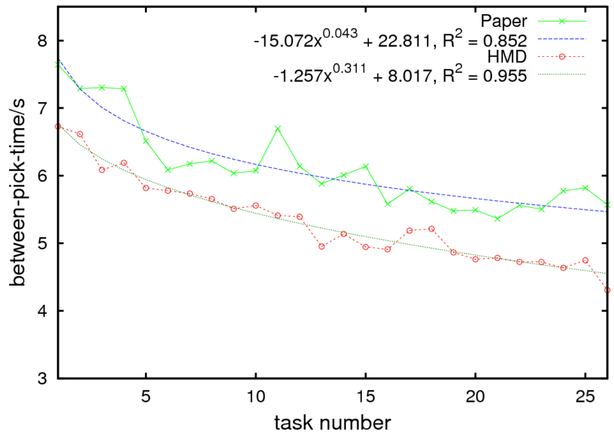

Figure 13: all picks

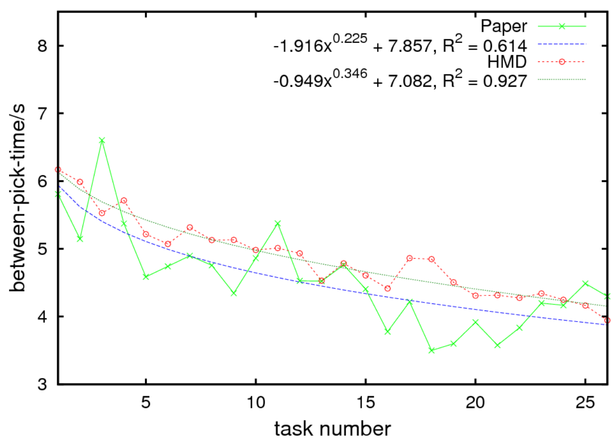

Figure 14: picks on same shelf

in industry. If examining the differences in particular hardware (e.g., two different HMDs) or control interfaces (e.g., speech versus context sensing), the experiment might be modified for better internal validity. For example, assemblages might be removed from the procedure and the same picking order might be given for all conditions.

Similar to Ockerman's findings [6], we found that our subjects performed the tasks in the order presented by the HMD and paper picking lists. Given that our subjects are novices, this behavior is to be expected. Perhaps true expert pickers may begin to optimize their own performance further with time. An *in-situ* study of current pickers in industry may be in order to see if the external validity of the study may be improved.

For our particular study, we plan to create better visualizations of the number of pieces and color/type of piece to be picked. We will also explore the role of eye dominance and hand dominance in picker's performance (for example, should the HMD be placed on the dominant eye?). Our HMD also had a relatively small exit pupil and no focus adjustment (the focus was fixed at arm's length), causing subjects difficulty in adjusting and using it. Perhaps a different HMD would have better results.

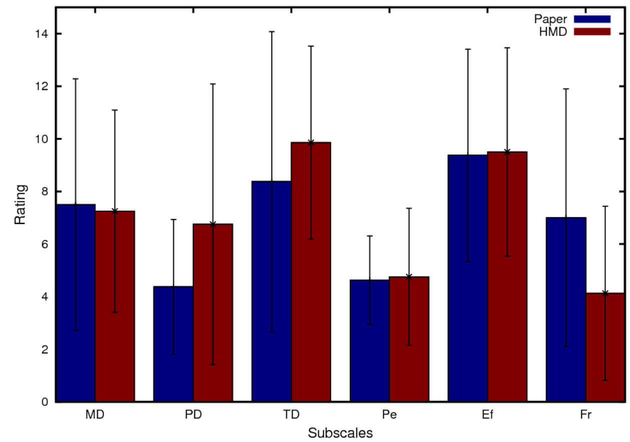

Figure 15: NASA TLX subscale ratings with bars indicating one standard deviation above and below the average (MD - mental demand, PD - physical demand, TD - temporal demand, Pe - performance, Ef - effort, Fr - frustration)

6.2 HMD vs. paper picking

HMD users made noticeable less mistakes where context awareness could be applied, and the total number of errors were trending to favor the HMD even though they were not statistically significant. After running the experiment, we realized that these numbers might be further improved by adding context "sensors" on the receiving bins to help determine if a part was transferred to the wrong bin. Thus, a piece in the wrong bin would count as a "context" error as opposed to an "other" error (as it is currently labeled). In the experiment, 75 error messages were displayed to the HMD user, and 59 were eventually removed due to the user's actions. Given this fact and the statistical results, context awareness seems to have a distinguishable effect in this experiment.

Note that the HMD results were hindered by wizard errors and possible "sensor" inconsistency. Real sensors, as described above, may result in better performance than that observed with our wizards.

Picking speed showed a distinct improvement with the HMD, especially as the users became more experienced. One thing that limited HMD speed is the relative emptiness of the bins. Subjects anecdotally reported that they found picking from (the relatively rare) full bins much easier than the mostly empty ones. Examining the picking speed on parts from the same inventory shelf reveals that most of the benefit may come from the ordering of the parts to better optimize the physical movement of the user. A paper picking list could be designed in a similar manner. However, the experiment did not consider the additional time required to print and retrieve a paper picking list between each task. In addition, a paper picking list can not be updated dynamically as with an HMD. A wirelessly connected wearable computer system may also be used to provide feedback to the inventory system as to empty or missing bins. The inventory system may then provide real-time advice for alternative locations for the required parts. The wearable system also allows better monitoring of the assembly process, possible helping to maintain ISO standards.

While the NASA TLX results did not show a statistically significant result at the end of the experiment, perhaps it would have showed a result if the survey was also given at the end of round one. Certainly, users had to spend effort in learning to use the HMD. While part of this effort may relate to the particular hardware chosen, any device would require accommodation by the user. Even so, the decrease in mistakes and increase in speed offered by context sensitivity would seem to offset the 2.5 hour training time required.

7. FUTURE WORK

We chose a paper-based picking list as a control because of its prevalence in industry today. An obvious future direction is to compare a speech-interface, similar to Vocollect's, with an HMD interface and a combination of the two. The organization of the display and the amount of picking context (which parts are next, how many parts remain on the list, etc.) could also be explored. We are also interested in repeating the study with different combinations of real sensors which may provide us with improved context-sensing (which seems key to improved accuracy and speed in this task). Finally, we would like to explore the effects of different HMDs, display and wearable mounting positions, and eye and hand dominance.

8. CONCLUSION

We have demonstrated an experimental procedure to evaluate the effect of wearable computers and contextual awareness for an inventory picking task. In our particular scenario, we have shown that the wearable system caused a significant reduction (versus paper-based picking list) in the amount of picking errors for the class of errors where context sensitivity could be applied. Wearable users were also faster pickers, due to the clustering of the parts into groups based on to which shelf they reside. We hope to continue to refine this experimental procedure and use it as a standard for exploring the effects of different wearable interfaces in the future.

Acknowledgment

This work is part of the project SiWear[1]. The project is funded with reference number 01MB07029A by the German Federal Ministry of Economics and Technology within the program SimoBIT [2].
We thank the Institute for Computer Science and Automation (IIA) at the University of Applied Sciences Bremen for their support at the user study.

[1]http://www.siwear.de/
[2]http://www.simobit.de/

9. REFERENCES

[1] J. Bartholdi and S. Hackmann. Warehouse and distribution science release 0.89. Technical report, Georgia Institute of Technology, January 2009.

[2] H. Byun and K. Cheverst. Utilizing context history to provide dynamic adaptations. *Applied Artificial Intelligence*, 18(6), 2004.

[3] J. Coyle, E. Bardi, and C. Langley. *The Management of Business Logistics: A Supply Chain Perspective*. South-Western College, Cincinnati, OH, 2002.

[4] S. G. Hart and L. E. Staveland. *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* Amsterdam: North Holland, 1988.

[5] D. Mizell. *Fundamentals of Wearable Computers and Augmented Reality*, chapter Boeing's wire bundle assembly project, pages 447–467. Lawrence Erlbaum & Associates, Philadelphia, PA, 2001.

[6] J. Ockerman. *Task guidance and procedure context: aiding workers in appropriate procedure following.* PhD thesis, Georgia Institute of Technology, Atlanta, GA USA, April 2000.

[7] D. Siewiorek, A. Smailagic, L. Bass, J. Siegel, R. Martin, and B. Bennington. Adtranz: a mobile computing system for maintenance and collaboration. In *International Symposium on Wearable Computers*, pages 25–32. IEEE Computer Society Press, 1998.

[8] D. Siewiorek, A. Smailagic, and T. Starner. *Application Design for Wearable Computing*. Morgan Claypool, San Rafael, CA, 2008.

[9] A. Smailagic, D. Siewiorek, R. Martin, and J. Stivoric. Very rapid prototyping of wearable computers: a case study of custom versus off-the-shelf design methodologies. *Design Automation for Embedded Systems*, 3(1):217–230, March 1998.

[10] T. E. Starner. Wearable computers: No longer science fiction. *IEEE Pervasive Computing*, 1(1):86–88, 2002.

[11] R. Stein, S. Ferrero, M. Hetfield, A. Quinn, and M. Krichever. Development of a commercially successful wearable data collection system. In *IEEE Intl. Symp. on Wearable Computers*. IEEE Computer Society, 1998.

[12] P. Yarin and H. Ishii. Touchcounters: Designing interactive electronic labels for physical containers. In *Conf. on Human Factors in Computing Systems (CHI)*. ACM Press, May 1999.