HNU Working Paper
Nr. 42

Fang Li

Oliver Kunze

Comparing pick-by-vision to pick-by-paper

An experimental assessment of pick times, error rates and user satisfaction

10 / 2020

Dr. Fang Li, Wissenschaftliche Mitarbeiterin für Innovative Logistik
Dr. Oliver Kunze, Professor für Logistik
Hochschule für angewandte Wissenschaften Neu-Ulm
University of Applied Sciences
Wileystraße 1, D-89231 Neu-Ulm

HNU HOCHSCHULE NEU-ULM UNIVERSITY OF APPLIED SCIENCES

Abstrakt

Wie wirkt sich Pick-by-Vision im Vergleich zu Pick-by-Paper (d.h. auf Basis einer herkömmlichen papierbasierten Kommissionierliste) auf die Picking-Zeit, die Picking-Fehler und die Benutzerzufriedenheit aus? Im Experiment konnten wir zeigen, dass die untersuchte Pick-by-Vision-Lösung die Fehlerrate in Bezug auf die Entnahme falscher Produkte dank der integrierten Anzeige der Lagerfach-Position im Blickfeld signifikant reduzieren konnte. Unsere Experimente zeigen auch, dass einige Kommissionierzeitkomponenten durch Pick-by-Vision positiv beeinflusst wurden, während andere Kommissionierzeitkomponenten negativ beeinflusst wurden. Außerdem schränken bestimmte technologische Probleme der getesteten Pick-by-Vision-Lösung derzeit die Benutzerzufriedenheit noch ein und könnten ein Hindernis für den Einsatz in einem industriellen Kontext darstellen.

Freie Schlagwörter: Augmented Reality, Mixed Reality, Order Picking, Pick-by-Vision, Pick-by-Paper

Abstract

This paper aims to investigate how pick-by-vision compared to a traditional paper-based pick list would affect the picking time, picking errors, and user satisfaction which are three most important factors for an industrial use. Results from experiments show that pick-by-vision would reduce the errors of picking wrong products significantly thanks to its integrated rack position display. Our experiments also show, that while some picking time components were positively affected by pick-by-vision, other picking time components were negatively affected. However, certain technological problems of pick-by-vision limits a better user-satisfaction and might be an obstacle for the use in an industrial context.

Keywords: Augmented Reality, Mixed Reality, Order Picking, Pick-by-Vision, Pick-by-Paper

JEL-Klassifikation: L81 , O14, O33

Table of Contents

1 Introduction

1.1 Background

Order picking is the process of finding and extracting required products from storage shelves to fulfill customer orders. It has been recognized that order picking takes an unreplaceable role among other warehouse activities in reducing logistics cost and improving customer satisfaction along the whole supply chain [1]. In addition, under the trend of product customization and fast market change, intra-logistic suppliers are challenged to improve their picking performance in warehouses for downstream partners [2].

Order picking aided by mixed reality (MR) / augmented reality (AR) is called “pick-by-vision” [3]. This innovative technology supports order picking with needed information projected at the side without sacrificing the view of reality [4]. According to the used software, pick-by-vision could be divided into two types: with and without position tracking system.

The simplest picking system without position tracking displayed equivalent information of a pick list by means of a small sentence written in the virtual view [5]. The typical system in the market e.g., Xpick from Ubimax company, provides graphical information to guide the user, based on a visual overlay chart where a picking grid is superimposed to represent the layout of shelving units with e.g., row color-coded and columns symbol-coded [6]. The bins to pick from are marked visually along with the picking number. The same concept could also be employed for order placing [7-9]. This picking system could be either implemented within wearable AR-glasses (e.g., Google Glasses) or by means of a cart-mounted display.

By a typical picking system with position tracking, cameras were installed in experimental warehouses to locate the user by following the marker equipped on the top of the user’s head mounted display (HMD) and tracking information was then transferred through WIFI to the wireless PC which is along worn by the user [3]. The user was guided to the rack position by a projected symbol (e.g., arrow, twisting tunnel, etc.) on the HMD. Picking time and mental stress were reported to be slightly reduced with pick-by-vision compared to paper-based pick list, but due to some technological problems pick-by-vision did not show advantages in reducing picking errors as expected. *(Funk et al., 2015)* proposed one mobile camera-projector cart which could detect picking and placing behavior through 3 pairs of top-installed depth cameras and projectors for in-situ projections to support order picking on the mobile cart *[10]*. Picking time, picking error and mental stress were reported to be the lowest among other picking methods.

Both mentioned tracking system required additional technical equipment – either users to be equipped with extra portable PCs or extra cameras to be installed in the warehouse. More developed Hardware such as Google Glasses and Microsoft Hololens become smaller, smarter, and slighter by integrating all necessary functions into the wearable end devices. (Hanson, Falkenström, und Miettinen, 2017) presented an AR application of Microsoft Hololens for kitting and tested it in a realistic laboratory experiment with five participants [11]. AR was competitive in terms of both time-efficiency and picking accuracy, both for single kit and batch preparation, compared to the traditional printed paper list. (Kretschmer et al., 2018) compared Microsoft Hololens with a conventional paper-based pick list and a tablet computer in the application of palletization [12]. From the study with 18 participants, no significant differences in terms of picking time and error rate were found between three alternatives whereas the usability of the tablet computer surpassed the Hololens and the pick list. (Lang et al., 2019) reviewed 14 literature sources addressing Microsoft Hololens’ application in production and logistics and found the most beneficial features of Hololens were recognized in enhanced ergonomics (i.e., free hands) and the internal tracking system [13]. However, an increased stress level was reported probably by manipulating the complicated control gesture and reading the small projection display.

1.2 Research Questions

The top level management question about picking can simply be stated as "Does it make sense to use pick-by-vision?" As this question is highly context sensitive (i.e. it depends on the situation of the company which does the picking), our research aims to evaluate one prototypic pick-by-vision solution in a defined experimental setting, and compare it to the simplest, non-technical picking solution – i.e. "pick-by-paper" (where the picker performs the picking based on a conventional paper printed pick list without any further ICT- support).

In this context, four research questions are going to be addressed:

1) *Picking time?* How would pick-by-vision affect the picking time, one of the most important factors of picking performance?
2) *Picking error?* In the aspect of identifying and reducing picking errors, traditional picking methods such as pick-by-paper and pick-by-voice usually require additional bar-code scanners or radio-frequency identification (RFID) scanners [14]. Another method is to implement sensible scales in mobile picking cart so that errors could be detected by checking the weight of picked items [15]. This paper is going to investigate in which way and how pick-by-vision could possibly affect the picking error rate, the second important factor of picking performance.
3) *User satisfaction?* Mental strain of the user delivers negative value with regard to the picking performance [16]. This paper analyses the user satisfaction (including but not limited to mental stress) of pick-by-vision as one additional factor for industrial employment.
4) *Technological obstacles?* The technological aspect is one of the success factors for industrial AR implementation projects [17]. This includes the processing power of HMD [10] and the interface design of smart glasses [18*]*. This paper is going to observe what kind of technological obstacles exist by the given system.

2 Experimental Design

The experiment was conducted in 2 phases – phase 1 (main experiment) and phase 2 (detailed experiment for picking time component analysis).

In each phase we compared the "pick-by-paper" scenario to the "pick-by-vision" scenario.

2.1 Experimental Design - Phase 1

In this experiment 28 different persons (so-called "pickers") performed a total of 100 picks each (50 with pick-by-paper and 50 with pick-by-vision). All participants are among 20-35 years old. 14 of them are university workers and the others are students. 6 participants are female. 2 participants have experiences with AR and 2 participants have experiences with order picking.

Picking Task

Pick order	Product ID	Name	Position	Nummber for picking	Picked?
1	513701	N 8X60/20 F (100)	B29	20	
2	652616	FPF-PZ 4,0X50 ZPF 200	A29	15	
3	18596	ADV BLANC - BTE DE 150 PCES	A28	20	
4	60510	ST 1 S8 S	A22	5	
5	507020	FBN II 8/10 K /RB 60	A27	4	
6	70005	SX 5X25 DUEBEL	A40	2	
7	50355	N 6X60/30 S (50)	B17	10	
8	514872	N 5X25/1 F (100)	B28	3	
9	657306	FPF ST 6,0X40 A4P 200	A31	7	

Figure 1 Experimental Picking Task (Excerpt)

The picking task for each round of our experiment in both scenarios was composed of 50 picking order lines (see Figure 1). To avoid the learn effect but make sure the results from two scenarios were completely comparable, the picking task employed by the pick-by-paper scenario included the same picking orders as in the pick-by-vision scenario but followed another sequence.

The picking task used in the pick-by-paper scenario was printed out on one A4 paper with a standard lay out.

The picking task used in the pick-by-vision scenario was transferred into the tested picking software based on the original picking data through two excel tables, one table for the storage including article ID, name, photo, and location of each item on both shelves and the other table for the picking task.

Picking Environment

The physical picking space was defined by two shelves (A and B) aligned in a 90° angle (see Figure 2). Each shelf was two meters tall and 1 meter wide. There were 52 “storage bins” stored on each shelf and 104 altogether. To identify each storage bin, respective labels (e.g., A42, B17) were pasted on the front side of the shelves. The items were stored in storage bins within shelves and were to be picked into “pick containers”.

Figure 2 Experimental Shelves’ Setting for Picking
(photographed by authors)

Different physical items to pick were sponsored by the company “fischerwerke GmbH & Co. KG” (in brief: Fischer). As the items for the experiment are mainly construction parts like screws and nails in different size, we stored them on the shelves according to the ABC method. That is, the most often used products are stored in the middle of the shelves while less used products are stored at the top and bottom of the shelves.

Mixed Reality Setting

The Microsoft Hololens 1st generation (MH1) integrates mobile computers, cameras, and projectors into one end device (see Figure 3). It is basically a wearable computer and by wearing this glasses, the user is provided with a computer generated superimposed view as the first virtual layer and the real environment as the basic view in the second layer. The superimposed view is projected to the glasses in front of the eyes of the user. To interact with the glasses, the user has to make different pre-defined finger gestures in front of the camera or speak to execute the ‘click’ and other functions.

To perform picking, MH1 had to be installed with a picking software which is not included in the default set of MH1. In our experiment we used the pre-commercial picking software, developed by a company which agreed to have this early prototypic version of the software tested by our university.

Figure 3 Microsoft Hololens 1st generation (photographed by authors)

In our tested software, a virtual space was firstly build up in accordance with the real physical shelves in X, Y, Z direction. To define the origin of co-ordinates, in our setting we had to scan one predesigned calibration paper/marker with MH1. This marker was placed on the ground (in our case in the bottom corner of both shelves) in order to calibrate MH1 and as basis for the co-ordinate calculus within the picking software.

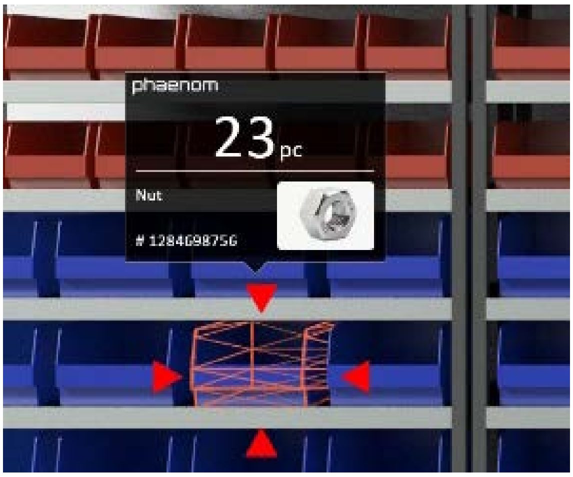

Figure 4 Augmented Reality User Interface

Source: www.phaenom.com [20]

Based on the internal position tracking system of MH1, the software indicated where the picker needed to go next by means of an "arrow" symbol on the MH1 display. Once the storage bin emerged in the view of the user, it was highlighted by means of a virtual pink lattice model overlay of the relevant storage bin on the MH1- display (see Figure 4 bottom). By firstly following the arrow, and secondly by the display of the lattice model which was virtually superimposed on the storage bin, the user was directed to the right storage bin. The picker was also informed about the right product (ID, name, and photo) and the number of items to be picked (see Figure 4 top) by means of an info box, which appeared in the display just above the lattice model. After picking the product, the user had to confirm the pick by means of a standardized finger gesture[1] (which was recognized by the MH1). The picker then got forwarded to the next pick position.

Organizational Setting

To simplify the organizational work of the experiment, each participant was invited to do both experiment scenarios directly one after the other: Scenario 1 – pick-by-paper und scenario 2 – pick-by-vision.

[1] An alternative option for pick-confirmation by means of voice control was not used in the experiment.

Eleven of the 50 picking orders required a "product set picking". Pickers in this situation had to pick the required number of parts of the set from the *same* mixed storage bin. For example, a product set consists of screws and nails and is stored in bin B48. The relevant picking order requires the picker to pick three sets (of screws and nails) from bin B48. For the pick-by-vision scenario, the picker was informed to pick the product set instead of picking one single product by means of a photo displaying a "product set". For the pick-by-paper scenario, the number of different parts in a set was pasted in front of the corresponding storage bin. For example, a "2" written on the bin-front indicated that there were two different item types within the mixed bin, which were needed for a set. This fact was explained to all pickers before their experiment.

For each picker, it took around 60 minutes to complete both experiment scenarios including explanation and training for MH1. Picking time and picking errors were recorded for each test. Picking errors referred to both number errors and product errors. For the scenario of pick-by-vision, each picker had around 5-10 minutes' time in advance to get himself trained with the finger gestures (needed for pick confirmation) by doing one turn of test picking. In this turn they only had to focus on how to make the right finger gesture.

Considering the similarity among products, it would take too long to re-sort every product after each round of the experiment if products of all 50 picked orders were dropped in one pick container. Therefore, this paper adopted the principle of 'one pick container for one picking order line'[2] for both scenarios. Therefore, the picker took one pick container for one picking order line and returned the filled bin to the control station which was one meter away from the shelves. Afterwards, the picker took another pick container from the control station to pick the next order line from the storage shelves. Meanwhile, the experiment-controller checked the picking results in time, returned items to their original storage bins on shelves, and recycled empty pick containers for the remainder of the test. By doing so, enough stock for the next experimental round was also guaranteed. The controller avoided coming into the way of the picker so that the picking time of the test was not affected. The controller did not intervene during the test unless there were technical problems (see section 4) to be fixed.

Setting for User Satisfaction Analysis

To investigate the user satisfaction of both picking scenarios, 4 indicators (see Table 1) were analyzed as follows: The last 8 participants among all 28 were asked to give a Likert-scale value between 0 and 100 for each indicator for both scenarios to show their perceived differences between pick-by-vision and pick-by-paper. 0 refers to quite easy/quite comfortable and 100 refers to quite hard/quite uncomfortable. A lower value means a better user satisfaction.

Table 1 User Satisfaction Indicators

Indicator	Explanation
Learnability	How easy is it to learn this picking method?
Readability	How comfortable is it to read the given information on the printed pick list/MR interface?
Operational Ease of Use	How easy is it to operate the picking technology?
Comfort of Use	How comfortable is it to wear this device? During experiment (30 minutes) / Long time of use (8 hours)

2.2 Experimental Design - Phase 2

As the result of phase 1 with respect to timely aspects was not as expected (see Figure 6), we tried to understand the timely results in greater details and conducted a follow up small scale experiment (= Phase 2).

[2] in contrast to the principle of 'one pick-container for all order lines of a pick list'

Overall Setting for Phase 2

Three extra participants were asked to do exactly the same experimental picking as the other 28 participants in phase 1. This time we filmed the pick operations, and based on the film time-code, we detected the different picking time components for a total of 150 picks (50 picks per participant) for each scenario. As one participant unintentionally skipped one order line, we had to remove the corresponding records in both scenarios, and thus the picking time composition was derived from a total of 149 test picks (see Figure 7).

Time Component Identification

We conceptually separated the different components (steps) of the picking times in accordance with VDI Richtlinie 3590 Blatt 1[3] [21] and added one initial step ("Grab") which was needed for our experiment[4].

Figure 5 shows the planned sequence of these steps for each scenario.

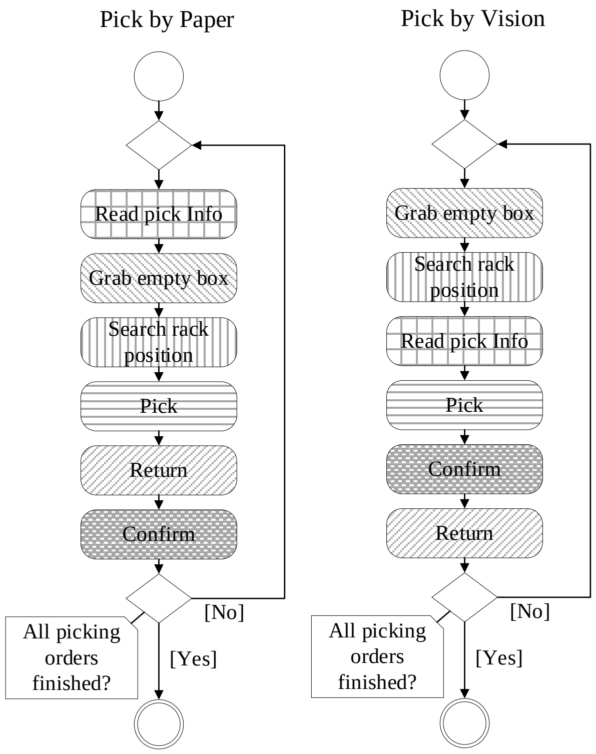

Figure 5 Comparison of the typical picking processes of two scenarios

For details on the time-component measurement in experiment 2 please refer to the methodological annex (chapter 9).

(Note, that during the experiment, some pickers deviated from the sequence of this plan due to their own preferences.)

[3] "Search rack position" = Bewegung des Kommissionierers zum Bereitstellungsort (VDI 3590 -1)
"Read pick info" = Vorgabe der Entnahmeinformation (VDI 3590 -1)
"Pick" = Entnahme der Artikel durch den Kommissionierer (VDI 3590 -1)
"Return" = Abgabe der Entnahme (VDI 3590 -1)
"Confirm" = Quittierung des Entnahmevorgangs bzw. der Entnahmevorgänge (VDI 3590 -1)
"Grab" = grabbing of empty pick container

[4] As we needed to check each pick for quality control, we asked the pickers to pick each pick position into a separate pick container – therefore before each pick this empty pick container had to be "grabbed" anew.

3 Experimental Results

Based on the above methodology, we obtained the following experimental results.

3.1 Average Picking Times (results from phase 1)

Experimental results showed that the pick-by-vision scenario was not faster than the pick-by-paper scenario as expected but conversely pickers with pick-by-vision used longer time (M=18.28 minutes, SD=2.80) to finish the same assignment than with pick-by-paper (M=16.2 minutes, SD=1.82) (see Figure 6).

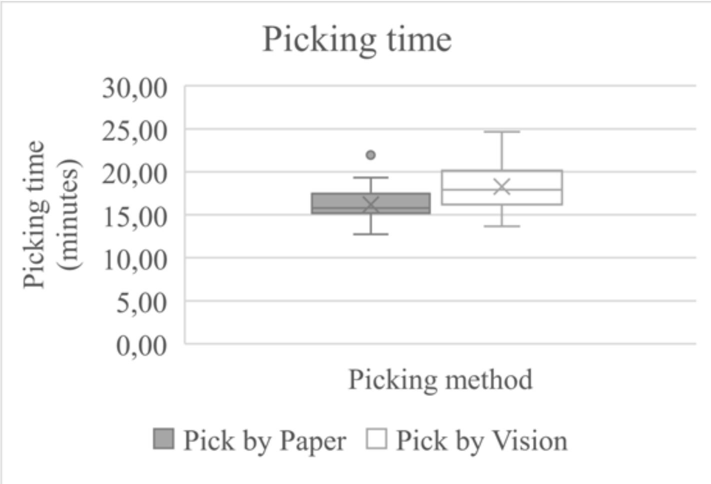

Figure 6 Picking time of the whole picking assignment of 50 picks each for both scenarios

As indicated in Figure 6, pickers in the tested pick-by-vision scenario needed 23.65 seconds in average to finish one order line while pickers in the tested pick-by-paper scenario needed 21.45 seconds. This unexpected result made us conduct experiment 2 in order to analyze the picking times in greater detail.

3.2 Picking Time Components (results from phase 2)

As depicted in Figure 7, the efficiency of the pick-by-paper scenario came mainly from the physical pick step with -1.9 seconds difference and the confirm step with -1.72 seconds difference in comparison with the pick-by-vision scenario. On the other hand, the pick-by-vision scenario showed better results in the read step with -1.27 seconds faster than the pick-by-paper scenario and in the search step with -0.61 seconds slightly faster. With regard to the return step, there was only a slight difference of 0.07 seconds between both scenarios.

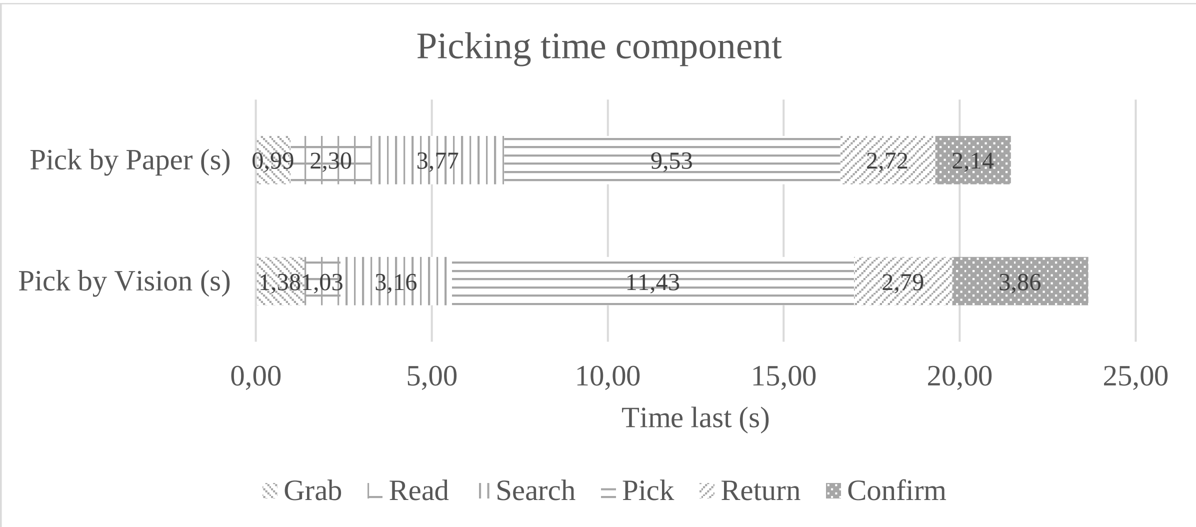

Figure 7 Picking Time Composition – averages of 149 test picks

This is how we interpret these figures.

Grab: The physical grabbing of the empty pick container took similar times in both scenarios. We speculate that wearing the MR-glasses slows down the movements of the picker, which would explain the 0.19s average time difference.

Read and Search: In the pick-by-paper scenario, pickers had to focus on two related numbers (i.e., picking number and rack position) in each order line among other text on the printed paper. Pickers needed time to find the information. Once they were forgotten, pickers had to go for a double check and this increased additionally the read time as a result.

Although the difference was slight by the search step, the pick-by-vision scenario did help pickers find the storage bin by the MR arrow and the lattice model. In the pick-by-paper scenario, if pickers got to a wrong storage bin due to mistake or false memory and found it afterwards, the resulted correction increased the search time. When the size or scale of shelves is getting larger, this advantage of the pick-by-vision scenario is likely to become even more relevant.

So in average the physical picking could start 1.49 seconds earlier in the pick by vision scenario.

Pick: Interestingly the physical picking took significantly longer in the pick-by-vision scenario, and overcompensated the previous time savings. Reasons for the big difference of time spent by the pick step are very likely related to the fact, that pickers in the post survey often mentioned that wearing the glasses made the real environment relatively dark and unclear.

Return: The physical return of the picked items took similar times in both scenarios. We speculate that wearing the MR-glasses slows down the movements of the picker, which would explain the 0.07s average time difference.

Confirm: Comparing to the easy tick with a pen in the pick-by-paper scenario, pickers often had two problems in making an effective confirmation in the pick-by-vision scenario. One was to execute the gesture in a way that it was properly recognized by the MH1. Although they had practiced this gesture in advance, this proved to be an obstacle from time to time. A second problem was related to the MR-display, which did not show the confirmation menu if the picker stood too close to the shelve. In this case the picker had to move one step backwards in order to see the MR confirmation prompt.

3.3 Picking Errors

The error rate is the percentage of the amount of wrong picking orders to the total amount of picking orders *(Schwerdtfeger et al., 2009*). Following this definition, the error rate of the pick-by-vision scenario (M=2.72%, SD=0.020) was slightly smaller than that of the pick-by-paper scenario (M=2.86%, SD=0.024) (see Figure 8).

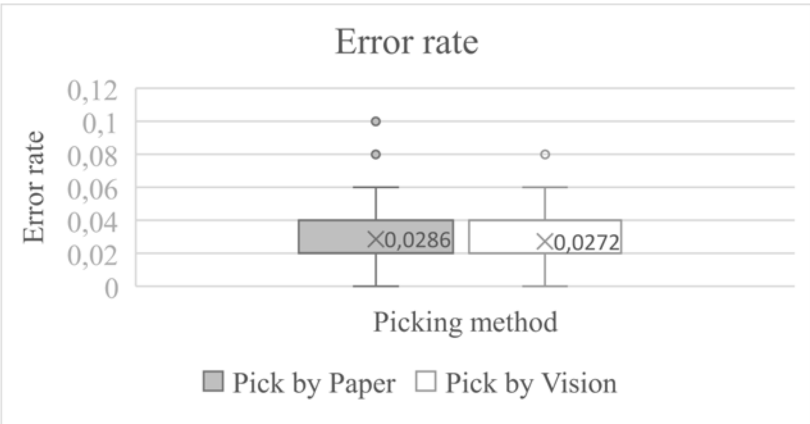

Figure 8. Error rate of both scenarios

Looking at the details, pickers in the pick-by-vision scenario altogether made 38 pick errors out of total 1395[5] picks while in the pick-by-paper scenario they made 40 pick errors.

As a picking error can either be a wrongly picked product, or a wrong number of items of the correct product (or a combination of both), we checked the nature of the picking errors in greater detail.

With regard to "**wrong product**" the pick-by-vision scenario performed significantly better (4 errors in total, i.e. 10 errors less than in the pick-by paper scenario).(see Figure 9).

Three errors of four resulted from the moved calibration where the MR lattice model located between two vertical neighbor bins on shelves (see section 4). Thus the pickers fetched the wrong storage bin without checking the provided photo on the MR interface. However, there was still one double pick in the pick-by-vision scenario produced by distraction. The picker forgot to confirm after returning the filled pick container as he was talking to others, and he picked again the already conducted order line.

The pick-by-vision scenario also reduced the risk of mixing shelve A and B to zero, and it happened 11 times all together out of 14 product errors in the pick-by-paper scenario. The rest three product errors in the pick-by-paper scenario were two vertical neighbor-mixing errors and one horizontal neighbor-mixing error. I.e. the pickers had the right rack position in mind but reached to the wrong neighboring storage bins by mistake. Still with regard to the "**wrong number of items**" the pick-by-vision scenario performed worse (34 errors in total, i.e. 8 errors more than in the pick-by paper scenario).

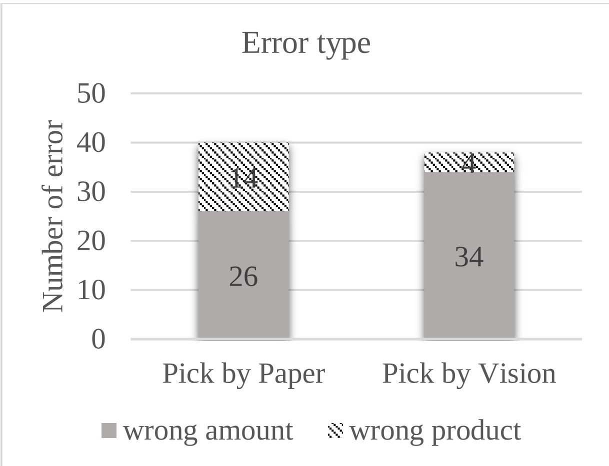

Figure 9. Error type and number of total 1395 picks for each scenario

Although neither principle could help to avoid counting mistakes, half pickers claimed that wearing MH1 made the view of the reality relatively dark and unclear so that it was possible to make counting errors. By reading information on the MR interface, six pickers told that they had problems in reading due to the too bright sunshine from the window in the experiment room.

3.4 User Satisfaction

To assess user satisfaction, we used a scale from 0 (quite easy/quite comfortable) to 100 (quite hard/quite uncomfortable). As shown from the post survey, the tested pick-by-vision scenario did not have any obvious advantages in the four predefined aspects. Results of the post survey are given in Figure 10.

In the aspect of *learnability*, the pick-by-vision scenario had a broad span of the major satisfaction value (SV) from 5 to 64 while the pick-by-paper scenario had quite concentrated values from 5 to 9. Averagely, it was more difficult to learn the pick-by-vision scenario (SV = 31) than to learn the pick-

[5] Due to a physical shelve replenishment error, error data for 5 picks needed to be deducted from the total of 1400 picks in each scenario.

by-paper scenario (SV = 7). As the pick-by-paper scenario did not require any special skills for any steps, it was the same friendly for all kind of users. But the question “How easy is it to learn the pick-by-vision scenario” depended a lot on users’ personal situation. The learnability in the pick-by-vision scenario indicated almost only the finger gesture for confirmation as this was the only step requiring interaction from pickers. There were pickers who got quickly used to this gesture and therefore they had no special preference of two scenarios in this aspect. There were also pickers having big difficulty in this step although they were given training in advance, and they gave the pick-by-vision scenario a much higher value than the pick-by-paper scenario as a result.

Two scenarios showed small difference in the aspect of *readability*. The pick-by-paper scenario had a slight advantage over the pick-by-vision scenario in the aspect of readability. The major SV given by pickers in the pick-by-vision scenario ranged from 9 to 43 (average SV = 27) and that in the pick-by-paper scenario lied in the range of 6-30 (average SV = 25). Noticeably, there was one picker who was not entertained at all in the read step in the pick-by-paper scenario.

With regard to *operational ease of use*, the SV for the pick-by-paper scenario ranged from 5 to 49 (average SV = 28) while the SV for the pick-by-vision ranged from 23 to 73 (average SV = 43). It is noticeable, that both scenarios had a broad span of SV. What was often criticized in the pick-by-paper scenario was no free hands and the need for a repeating information check on the pick list where a positive feedback was reported in the pick-by-vision scenario with regard to these two aspects.

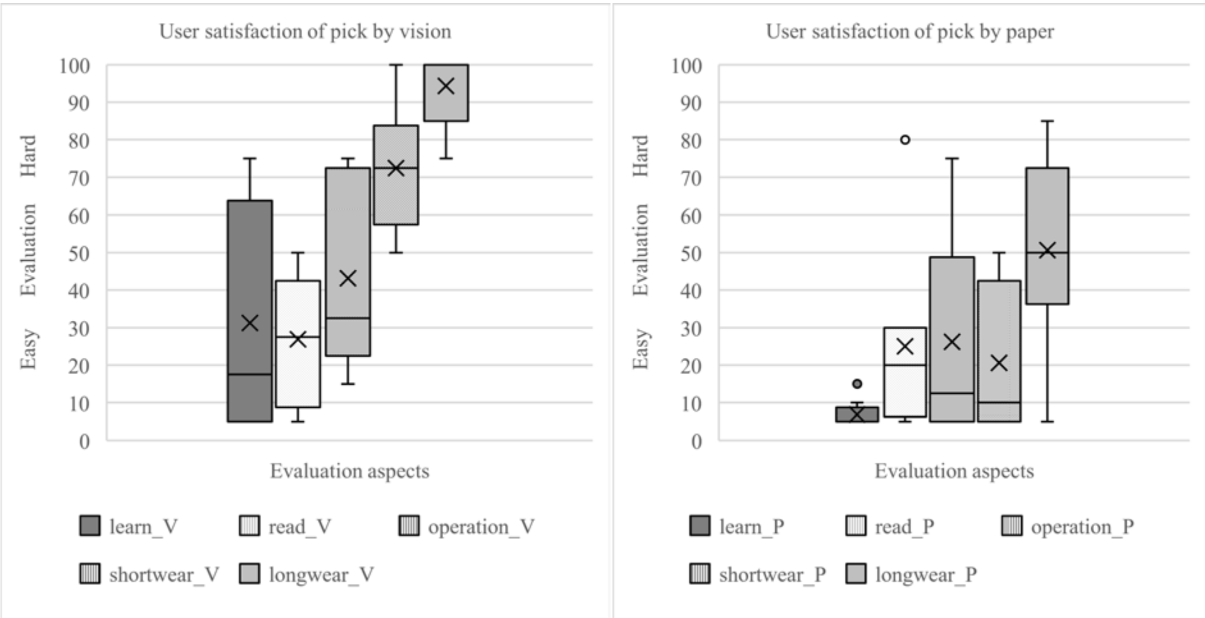

Figure 10. User satisfaction of pick-by-paper/vision

The evaluation for the *comfort of use* of the tested pick-by-visions scenario was the worst among all four factors as well for the short experimental period (average SV = 73) as for the speculation of a long industrial use (average SV = 95). The glasses tested in the experiment weighed 579 grams. Some pickers felt already dizzy after 30 minutes of use in the experiment and all of them imagined that wearing these glasses for 8 hours to work in a real warehouse environment was not desirable. However, working with the pick-by-paper scenario was neither evaluated as ideal for a long time picking (average SV = 51).

When asked the preference between these two scenarios, half respondents inclined to the pick-by-vision scenario while the other half preferred to the pick-by-paper scenario.

4 Technical Observations & Room for Improvement

As the software tested in the experiment (phase 1 and 2) was not yet a commercial solution for picking, our observations as well as results from the experiment remain only intermediate outcome. We introduce our observations of encountered technical problems and expect new versions of the picking software soon to come.

As observed, 2/3 of the participants encountered at least one of the following technical problems during the use of the pick-by-vision software in the experiment:

1) Moved calibration – The calibration was not fixed after the first scanning of the marker; however, it moved slightly or drastically. For example, the MR lattice model was located sometimes between two horizontal neighbor bins or two vertical neighbor bins in reality. When pickers wearing the Glasses looked down to the ground for a while during picking/counting, the virtual interface could be projected on the ground instead of on the original shelves. Pickers in both cases had to scan the marker again so as to come back to the original calibration. This unexpected problem could possibly be resolved by the updates of the operation system of the Glasses which did not any more support all existing functions of the picking program, as speculated by the software developer. Required solutions were not yet found.
2) Default mode of the picking program – Pickers saw only a small white point and nothing else when the program moved into a default mode for whichever reason. After doing a "click" gesture, it showed on the interface "tap here to start" and pickers must follow this structure in order to go back to the last picking status. As consulted by the software developer, the reason could be either unintentional finger gestures received by the Glasses or relatively dark light in the experiment room.
3) Skipping of unfinished picking orders – Some pickers had difficulties with practicing the confirmation gesture. If pickers made the confirmation gesture by accident during their movements (but not at the confirmation step) the current picking order was skipped (i.e. it was recorded as finished, and the subsequent picking order was initiated). Once this happened, it was impossible to go back.
4) Battery exhaustion - the MH1 used in the experiment did only last for a continuous use of around three hours after a full charging of the inserted-battery during an eight hours charging period.

5 Selected Insights from a Management Perspective

Besides the results from the experiment, we want to share our assessment of the pick-by-vision solution from a management perspective. We thus present an ad-hoc qualitative management decision scale of + / 0 / - / ? in Table 2. (+: positive aspect; 0: indifferent aspect, -: negative aspect, ?: cannot be assessed in the given context.)

Table 2 Insights from a management aspect

	Pick-by-vision	Comments
initial investment	?	MR glasses & picking software
time cost	-	small additional picking effort (9% in our experiment)
error cost	+	significant reduction of "wrong product picked"
language requirement	+	no requirement of special language knowledge due to picture & figure display
user satisfaction	-	human interface design (incl. physical stress of wearing glasses) – room for improvement with regard to our experimental setting
flexibility & scalability	+	ease to adopt method in a changing environment (no inbuilt hardware)
battery charging times	-	long battery charging times for max 3h of productive use – room for improvement with regard to our experimental setting

As an **initial investment**, pick-by-vision requires MR glasses and a picking software. Prices for one piece of glasses range in the lower 4-digit US$ price range (e.g. the second generation of Microsoft Hololens (MH2) glasses today cost 3.752 Euro incl. VAT per piece [22], and the second generation of google glasses (Glass Enterprise Edition 2) is offered starting from 1.000 US$ per piece [23]). Costs for picking software need to be added and depend on software provider, license model and optional volume discounts.

With regard to **time costs** our experiment showed a quicker start of the physical picking per pick position due to less orientation time, but this advantage was overcompensated by a higher effort for physical picking and especially for pick confirmation. Therefore, future pick-by-vision solutions might profit from alternative technical options for pick confirmation (e.g. by portable tactile confirmation devices or by voice control) on the one hand, and from a less shaded real world view through the lenses on the other hand.

In terms of **error costs,** we see a clear advantage in pick-by vision solution with regard to the picks of "wrong products". With regard to the pick of "wrong number of items", an additional check technology (as e.g. scales for weight check) complementing the solution seems advisable from our point of view.

Regarding the integration of foreign workers with **language** problems into a picking workforce, we see an advantage in the pick-by-vision solution, as pictures, pointers and figures are sufficient for guiding the picker to do the relevant picking tasks.

With regard to the aspect of **user satisfaction**, we want to stress, that we tested one pre-commercial test-version in combination with one specific hardware for glasses (and not several commercial productive versions of different providers), only. Therefore, our results on user satisfaction cannot easily be generalized. Still, what we tested was a promising starting point, and we speculate that the detected issues will be improved in the near to mid-term future.

With regard to **flexibility** & scalability we see a high potential for pick-by-vision from a managerial perspective, as no ICT-infrastructure needs to be added to the shelves (as e.g. in a pick-to-light solution), and the solution can be adopted to a changed warehouse layout relatively easily.

Finally, we see room for improvement in the **battery re-charging time** cycles – a battery recharging cycle should at least span the picking time between two regular breaks, and if possible, a whole shift-duration of a picker.

6 Conclusion

In contrast to our expectations from previous pre-tests, we couldn't prove a positive impact of pick-by-vision on total pick times. Still we could show how pick-by-vision speeds up the orientation times for pickers. We also found a positive impact of pick-by-vision on pricking error rates with regard to "wrong products".

Still, we want to critically discuss our methodology and our results in greater detail.

6.1 Critical Assessment of Methodology

From a retrospect perspective, we see room for improvement in our experimental design as follows.

Scale of survey – we made a comparative survey based on a total of 1.395 pick in each scenario and 28 persons in total. As always a larger number of picks and persons would have resulted in better statistical quality.

Physical dimension of survey – we used a limited shelf-design of two shelves, only. Therefore, we cannot be sure, if the positive effects of pick-by-vision were larger with regard to orientation time components, provided that physical spaces were larger, too.

Comparative basis – we compared just one technical (pre-commercial) pick-by-vision solution with one non-technical pick-by-paper solution. For future experiments it would be desirable to compare different commercial versions of pick-by-vision solutions with other picking solutions (e.g. pick-to-light, pick-by-voice).

Time component analysis – the concept of picking time components was missing in our initial experimental design. Thus we needed to add phase 2 of our experiment. But even so it was difficult to measure the different time components (see Appendix). A more precise instruction for the workers on how they should perform their picking tasks would have reduced this problem, but it would have rendered this “instruction” into a precondition for the experiment – i.e. different sets of instructions could have come to different experimental results.

Bias factors – we see the following set of possible bias factors, which were not (or not fully) covered by our experimental design:

- Familiarity with pick-by-vision solution – even though we introduced a short “learning phase” prior to our experiment, we cannot rule out that a longer learning phase would have generated different results – especially the unfamiliarity with gesture control for pick confirmation may have had a bias effect on our results
- Obstacles caused by software and hardware (especially offset in shelve display, process interruptions and battery failure – which were detected after the experimental design, only, and were observed during the experimental execution) – these aspects were considered in the analysis of the results – affected data were eliminated.
- Mistakes in physical shelve replenishment – these aspects were considered in the analysis of the results – affected data were eliminated.
- Quality discipline – as we had neither set specific instructions nor incentives for the participants of the study with regard to picking quality, we cannot rule out, that some participants just did not sufficiently try to avoid picking mistakes, and thus focus more on getting finished with the job.
- Other uncontrolled variables were light & illumination situation, psychological disposition of pickers (incl. daytime specific periods of fatigue), plus (as always) unknown unknowns.

6.2 Critical Assessment of Results

The quantitative results of our experiments are related to our experimental setting, but within this setting, they are precise. The qualitative results (interpretation of results) are based on experimental observations and thus seem robust from our point of view. The qualitative assessments from a management perspective (see chapter 5) are founded on personal experience and opinions, and thus include potential for speculation.

6.3 Need for Further Research

We see a need for further research with regard to the analysis of newer versions of hardware (2nd generation of Microsoft Hololens) and software, larger scale experiments (larger physical space as well as more pick positions) and with regard to pick time component identification in heterogeneous pick execution variants (i.e. different details in picking process execution based on the preferences of each individual picker).

6.4 Outlook

We assume that the technology of mixed reality MR (or augmented reality AR) will develop further in the near future. We see potential in this technology especially for logistics operations incl. picking, assembly and maintenance works. Further experimental research on the quantification of these potential benefits will help to detect technical shortcomings as well as to detect commercial potentials of MR (or AR).

7 Acknowledgements

This paper is fully supported by Innovation Transfer for the Region Donau-Iller-Riss project (InnoSÜD) (2018-2022). The authors would like to thank Phaenom GmbH for providing their pre-commercial solution and technical support and Fisher Company for providing picking items. Appreciation is also given to Mr. Walid Taha for sharing his first experimental results and all experiment participants for their support and feedback. Last but not least, authors would like to thank Mr. Johannes Som at the university of applied science in Ulm in Germany for introducing different picking technologies (i.e., pick by voice, pick by light, pick-by-vision with cart-mounted display/Google Glasses with conjunction of RFID finger scanner) in the logistic laboratory and providing the first hands-on experiences.

8 References

1. Bartholdi III, J. J., & Hackman, S. T. (2011). Warehouse and distribution science. Supply Chain and Logistics Institute, School of Industrial and Systems Engineering, Georgia Institute of Technology.
2. De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. European journal of operational research, 182(2), 481-501.
3. Schwerdtfeger, B., Reif, R., Gunthner, W. A., Klinker, G., Hamacher, D., Schega, L., ... & Tumler, J. (2009, October). Pick-by-Vision: A first stress test. In 2009 8th IEEE International Symposium on Mixed and Augmented Reality (pp. 115-124). IEEE.
4. Reif, R., Günthner, W. A., Schwerdtfeger, B., & Klinker, G. (2010, March). Evaluation of an augmented reality supported picking system under practical conditions. In Computer Graphics Forum (Vol. 29, No. 1, pp. 2-12). Oxford, UK: Blackwell Publishing Ltd.
5. Stoltz, M. H., Giannikas, V., McFarlane, D., Strachan, J., Um, J., & Srinivasan, R. (2017). Augmented reality in warehouse operations: opportunities and barriers. IFAC-PapersOnLine, 50(1), 12979-12984.
6. Ubimax homepage, https://www.ubimax.com/frontline/xpick, accessed 9th of August 2020.
7. Battini, D., Calzavara, M., Persona, A., & Sgarbossa, F. (2015). A comparative analysis of different paperless picking systems. Industrial Management & Data Systems, 115 (3), 483-503.
8. Guo, A., Raghu, S., Xie, X., Ismail, S., Luo, X., Simoneau, J., ... & Starner, T. (2014, September). A comparison of order picking assisted by head-up display (HUD), cart-mounted display (CMD), light, and paper pick list. In Proceedings of the 2014 ACM International Symposium on Wearable Computers (pp. 71-78).
9. Weaver, K. A., Baumann, H., Starner, T., Iben, H., & Lawo, M. (2010, April). An empirical task analysis of warehouse order picking using head-mounted displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1695-1704).
10. Funk, M., Shirazi, A. S., Mayer, S., Lischke, L., & Schmidt, A. (2015, September). Pick from here! An interactive mobile cart using in-situ projection for order picking. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 601-609).
11. Hanson, R., Falkenström, W., & Miettinen, M. (2017). Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. Computers & Industrial Engineering, 113, 570-575.
12. Kretschmer, V., Plewan, T., Rinkenauer, G., & Maettig, B. (2018, January). Smart palletisation: Cognitive ergonomics in augmented reality based palletising. In International Conference on Intelligent Human Systems Integration (pp. 355-360). Springer, Cham.
13. Lang, S., Kota, M. S. S. D., Weigert, D., & Behrendt, F. (2019). Mixed reality in production and logistics: Discussing the application potentials of Microsoft HoloLensTM. Procedia Computer Science, 149, 118-129.

14. Thomas, C., Panagiotopoulos, T., Kotipalli, P., Haynes, M., & Starner, T. (2018, October). RF-pick: comparing order picking using a HUD with wearable RFID verification to traditional pick methods. In Proceedings of the 2018 ACM International Symposium on Wearable Computers (pp. 168-175).
15. Wu, X., Haynes, M., Guo, A., & Starner, T. (2016, September). A comparison of order picking methods augmented with weight checking error detection. In Proceedings of the 2016 ACM International Symposium on Wearable Computers (pp. 144-147).
16. Haase, J., & Beimborn, D. (2017, June). Acceptance of Warehouse Picking Systems: A Literature Review. In Proceedings of the 2017 ACM SIGMIS Conference on Computers and People Research (pp. 53-60).
17. Masood, T., & Egger, J. (2019). Augmented reality in support of Industry 4.0—Implementation challenges and success factors. Robotics and Computer-Integrated Manufacturing, 58, 181-195.
18. Kim, S., Nussbaum, M. A., & Gabbard, J. L. (2016). Augmented reality "smart glasseses" in the workplace: industry perspectives and challenges for worker safety and health. IIE transactions on occupational ergonomics and human factors, 4(4), 253-258.
19. Taha, W., 2017. Verbesserung der Kommissionierleistung und -qualität durch Einsatz von Augmented Reality in der Logistik und in der Medikamentenbereitstellung. Bachelor thesis at the applied-science university Neu-Ulm, submitted on Aug 4th, 2017.
20. Phaenom homepage, www.phaenom.com, accessed 9th of August 2020.
21. VDI-Fachbereich Technische Logistik (1994). VDI manual Technical Logistics - Volume 8: Material Flow II (Organisation/Control). VDI 3590 Blatt 1. VDI-Gesellschaft Produktion und Logistik.
22. Microsoft homepage, https://www.microsoft.com/de-de/hololens, accessed 7th of August 2020.
23. Computerwoche homepage, https://www.computerwoche.de/a/google-glass-enterprise-edition-2-ab-sofort-fuer-alle-erhaeltlich,3548391, accessed 7th of August 2020.

9 Appendix – Methodology Details for Experiment 2

First of all, three different ways were observed from pickers in experiment-phase 2 to do their job.

a) Pickers brought the empty pick container from the control station to shelves for picking.
b) Pickers took the storage bin from shelves to the control station and move items to the pick container there.
c) Pickers picked items directly to hand at shelves and put back to the pick container at the control station.

Pickers switched from a) to b) for the items stored at the top and bottom of shelves.

As depicted in Figure 5 we conceptually sequenced the time components according to the type a) because a majority of pickers preferred to follow this typical process. In type b) and c), some steps (time components) exchange their sequence in both scenarios. For example, in type b) the step grab and search exchange in the pick-by-paper scenario and in type c) the grab step was skipped in both scenarios. Other differentiation from the typical picking processes of both scenarios are going to be clarified in the following steps.

The picking time components were identified with each start event (⇤), include behavior (↔), end event (⇥), and respective problems in the time component delimitation & resolution (‽) as followed in each step.

Grab: This step is described according to the abovementioned three different pick ways. As pickers skipped the grab step in type c), the time for grabbing was defined as zero.

- Pick-by-paper

	a)	b)	c)
⇤	Move eye sight away from reading	Touch the storage bin after searching	/
↔	Reach out hand, fetch the pick container	Take out the storage bin, walk back to the control station, and pull out one empty pick container from the stack if needed	/
⇥	Turn to shelves for searching	Start picking	/
‽	?: Pickers often grabbed the pick container while still kept on reading. !: 1 second was given to the grab step.	/	/

- Pick-by-vision

	a)	b)	c)
⇤	Finishing returning the filled pick container	Touch the storage bin after searching	/
↔	Reach out hand, fetch the empty pick container	Take out the storage bin, walk back to the control station, and pull out one empty pick container from the stack if needed	/
⇥	Turn to shelves for searching	Start picking	/
‽	?: Pickers often grabbed the pick container while they searched. !: 1 second was given to the grab step.	/	/

***Read**:* It is reading and remembering needed information (i.e., picking number and rack position). It was observed this step was often overlapped with other steps (i.e., grab, search).

- Pick-by-paper

	a)	b)	c)
⇤	Drop down the pen after confirmation on printed paper		
↔	Look at paper, read, and remember needed information		
⇥	Reach out hand to grab pick container	Turn to shelves for searching	
‽	?: Pickers sometimes read again during the search and pick step. !: The time spent on sub-sequential reading was separated out and added into the read step.		

- Pick-by-vision

	a)	b)	c)
⇤	Touch the storage bin after searching		
↔	Look at the MR interface		
⇥	Start picking		
‽	?: Pickers usually took a glance quickly at the information shown on the MR interface without special time spent, as they had only to look at the picking number. !: 0.5 second was allocated to the read step unless obvious reading behavior was observed. Pickers sometimes read again during the pick step. !: The time was separated out and added into the read step.		

***Search**:* It is the time for searching until the storage bin is located. Errors of picking wrong items might be produced in this step when a wrong storage bin is located.

- Pick-by-paper

	a)	b)	c)
⇤	Face to shelves after grabbing	Face to shelves after reading	
↔	Move eye sight, walk to shelves		
⇥	Touch the storage bin		
‽	?: Pickers realized that they were picking/have picked wrong items from wrong storage bins and therefore corrected their picking. !: The time spent on picking wrong products and the time for correcting (i.e., put back wrong items until the right storage bin was touched) were allocated to the search step.		

- Pick-by-vision

	a)	b)	c)
⇤	Face to shelves after grabbing	Successful confirmation	
↔	Move eye sight, walk to shelves		
⇥	Touch the storage bin		
‽	?: Pickers were sometimes confused if they reach the right storage bin because the MR lattice model did not overlay precisely of related storage bin. !: Pickers were told in this case to calibrate Hololens again by looking at the marker with Hololens on the ground and this time was deducted from final results as it was long and not expected to happen with other mature software. !: Otherwise, pickers were told to check the MR photo of relevant items and this time was allocated to the search step.		

***Pick**:* This is the essential step of the whole picking processes where counting errors and product errors of product set-picking might be taken place. This step is described as the same for both scenarios according to the three different grab ways.

- Pick-by-paper / vision

	a)	b)	c)
⇤	Touch the storage bin	Touch items in the storage bin	Touch the storage bin
↔	Pull out the storage bin, move/count required number of items from the storage bin to the pick container	Move/count required number of items from the storage bin to the pick container at the control station, move the filled pick container close to the controller if applied	Pull out the storage bin, move/count required number of items from the storage bin to hand
⇥	Push the storage bin back	Turn to shelves for returning	Push the storage bin back
‽	?: Pickers realized afterwards they had picked wrong number of items or made errors by missing items of product set. !: The time for correction was counted to the pick step.		

***Return**:* This step is described according to the three different pick ways. In the pick-by-vision scenario, pickers returned firstly before they confirmed by the pick way b) and c).

- Pick-by-paper

	a)	b)	c)
⇤	Complete pushing back the storage bin	Turn to shelves after picking	Complete pushing back the storage bin
↔	Turn body from shelves, walk to the control station, drop down the filled pick	Walk to shelves, put back the storage bin to its original	Turn body from shelves, walk to the control station, pull out one empty pick container from the stack if needed, drop

	container, and move it close to the controller if applied	position, walk back to the control station	down items in the pick container, move the filled pick container close to the controller if applied
⇥	Start reaching the confirmation pen		
‽	/	/	?: Pickers had to return more than one time during the pick step if the amount of picked items were too large for one hand to hold (e.g., >10). !: The time spent on multi-return was separated from pick and allocated to the return step.

- Pick-by-vision

	a)	b)	c)
⇤	Successful confirmation (voice from MH1)	Turn to shelves after picking	Complete pushing back the storage bin
↔	Walk to the control station, drop down the filled pick container, and move it close to the controller if applied	Walk to shelves, put back the storage bin to its original position	As same as the pick-by-paper scenario
⇥	Reach out hand for grabbing	Walk backwards for confirming	
‽	/	/	As same as the pick-by-paper scenario

Confirm*:* Pickers had to confirm in different ways in two scenarios.

- Pick-by-paper

	a)	b)	c)
⇤	Reach the confirmation pen		
↔	Grab pen in hand, proof picked information if applied, tick in relevant blank on the printed paper		
⇥	Drop the pen		
‽	?: Pickers sometimes read the information once again before they ticked. !: This time was counted to the confirm step. Once the pen was dropped down, the time was counted to the read step when pickers looked on at the paper. Pickers read at the same time as they did confirmation. !: 2 seconds was deducted and allocated to the read step.		

- Pick-by-vision

	a)	b)	c)
⇤	Re-looking at the MR interface after picking	Finish putting back the storage bin	Finish returning picked items in the pick container
↔	Move backwards to a proper position, lift up one hand, make figure gesture		
⇥	Voice of the successful confirmation is heard		
‽	?: Pickers had to move around 1 meter backwards in order to make the confirmation gesture and they returned the pick container at the same time as they walked backwards. !: 1 second was deducted and allocated to the return step.	/	/

In addition, there were some other problems observed at any step during the whole picking process. Problems as well as solutions are described as follows:

- Pickers sometimes needed to re-adjust the Glasses (e.g., the Glasses slapped down) during the picking process. Time for the adjustment was counted to the relevant step where it happened.
- When the pick-by-vision software needed to be reset (default mode), the time spent on the reset was deducted.