

# Supporting Order Picking with Augmented Reality

Björn Schwerdtfeger\*

Gudrun Klinker†

Technische Universität München  
Fakultät für Informatik  
Boltzmannstraße 3  
85748 Garching b. München, Germany  
<http://campar.in.tum.de>



Figure 1: Two different Augmented Reality visualizations for supporting the process of order picking. It shows images out of moving sequences. a+b) A Bezier curve based Tunnel guiding the users attention to a box from where to pick items. The tunnel fades to transparent when the actual box is in the FoV. The actual box is highlighted by an opaque Frame. c+d) An arrow extended with a rubberband guides the users view to the actual augmentation, which is again a Frame.

## ABSTRACT

We report on recent progress in the iterative process of exploring, evaluating and refining Augmented Reality based presentation schemes for logistic applications. In this context we have evaluated different ways of presenting working instructions with HMD-based Augmented Reality systems to support the order picking process. Order picking means that workers have to pick items out of numbered boxes, according to a work order. For small boxes, the requirements for having precise and clear visualizations are demanding. In this paper we report on our findings from three user studies and from presenting the system at several exhibitions. They led to several optimizations according to the visualization schemes and the selection of a suitable HMD. Finally we present the resulting setup, which consists of combined visualizations to precisely and efficiently guide the user, even if the augmentation is not always in the Field-of-View of the HMD.

**Index Terms:** H.5.1 [ INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities; Evaluation/methodology; H.5.2 [ INFORMATION INTERFACES AND PRESENTATION]: User Interfaces —User-centered design;

## 1 INTRODUCTION

In the context of *supra-adaptive*<sup>1</sup> logistics applications [?], there is a need for highly efficient and intuitive ways to present information. The base of supra-adaptive systems are flexible workers. They have to adapt to new working conditions and environments quickly, frequently, and with minimal training. On this account they always need to be provided with the right information at the right time.

\*e-mail: schwerdt@in.tum.de

†e-mail:klinker@in.tum.de

Strictly speaking they need detailed working instructions, which have to be presented highly intuitively and exactly. This enables the workers to immediately start an efficient and error-free execution of an arbitrary job.

Our use case in this context is the order picking process. We have iteratively developed and evaluated several mobile Augmented Reality based visualizations to support this process efficiently. This paper reports on three empirical user studies that we have conducted to develop and improve our system. Overall we have gathered data from 64 subjects (18 in Test 1, 34 in Test 2, 14 in Test 3) picking altogether 5080 items (1620 in Test 1, 1940 in Test 2, 1512 in Test 3) out of different boxes in a warehouse, guided by one of our various Augmented Reality visualizations. Furthermore we have presented the system several times at public events, which we also used to note people’s reactions and observing how they behave with the different visualizations. This iterative process helped us improve our visualizations to bring them to an efficient level.

Before presenting our three user studies, we discuss the challenges of designing systems to support the order picking process, and what this means in the scope of Augmented Reality. We present the system setup we used.

### 1.1 Order Picking in Logistics Applications

In order picking tasks, workers collect sets of items from an assortment in a warehouse according to a work order and deliver them to the next station in a precisely designed material flow process [?]. The efficiency of such picking processes is divided into time measurements of four interleaved tasks: the *base time* for getting the next order information, the *dead time* during which a worker interprets and understands the order as a 3D navigation and picking task, the *way time* during which the user physically moves to the selected item and the *picking time* to actually grab the item. While the base time, way time and picking time have been subject to many process optimizations, we focus on providing visualizations to reduce the dead time. However, it is not only important to optimize this process in time, but also to prevent users from making errors. Wrongly picked items can lead to high follow-up costs. For this reason we

<sup>1</sup>to be able to adapt with minimal effort to global dynamic changes

have to ensure that the worker picks out of the right box with high reliability. Which is a challenging task, as according to Gudehus[?], there exists no such zero error picking system. In current processes errors in order picking systems are reduced by follow-up checks like barcode scanning or checkweighing.

Traditionally, order picking is accomplished by providing workers with printed-out pick lists of articles, stating their position in the warehouse, the amount to be collected, and short descriptions. To increase efficiency and to reduce the number of picking errors, current industrial setups start using two techniques [?]: *Pick-by-Light* approaches use lamps (LEDs) that are installed with each lot of items on the warehouse shelves. When a new order is processed, lamps corresponding to each item on the list are turned on to show the lot and the amount to pick. Such a system cannot simultaneously support multiple workers picking items in the same shelf area. In *Pick-by-Voice systems*, workers wear headsets and receive auditory information about the next item to collect. Yet, the system requires considerable mental effort since workers have to remember what has been said, and they have to relate it to the geometrical layout of the area.

## 1.2 Supporting Order Picking with Augmented Reality

When designing a system to support the order picking process we basically have to support both phases of the navigation (pathfinding and picking). At first we have to guide users the path to the right shelf, which is followed by a another visualization for the actual picking process. As we are dealing with HMDs and the problem of a small field of view, we must support the actual visualizations by a meta visualization, to help users to find the augmentation.

## 1.3 Related Work

The task of guiding the user in a warehouse to specific box can be seen as an abstraction of the work by Biocca et al[?]. They basically try to cue visual attention to any physical or virtual object in 4 steradians using mobile Augmented Reality systems. They emphasize the challenge to minimize the workload by using attentive interfaces. Whereas those interfaces dynamically prioritize the information they present to their users, such that information processing resources of both user and system are optimally distributed across a set of tasks [?].

Feiner et. al[?] developed a rubber band like visualization leading to an on-screen or off-screen object, combined with highlighting the object, to help the user find it when wearing a tracked HMD.

More systems to overcome the problems with large and complex warehouses as well as with inexperienced workers, have been discussed in [?].

## 1.4 System setup

As we needed our system to be flexible and adaptable to several scenarios and evaluation setups, we based it on a distributed component framework. Originally this was the DWARF middleware [?], which we later exchanged by a specially developed framework. This was mainly due to new developments in the area of tracking middleware. The new Framework is Java-based and uses TCP/IP communications between components. This allows us to easily deploy the system in different scenarios and interchange visualization components for our evaluations. To make the system independent of a dedicated tracking system, we integrated the Ubitrack middleware [?] in our system (using the Ubitrack Java-bindings). Ubitrack allows us to change the tracking setup by simply changing the configuration file (UTQL) using a graphical editor [?]. For the current evaluations, we mainly rely on the ART Dtrack infrared optical camera tracking system, due to its availability in our lab. The ART Dtrack system requires that retro reflective marker bodies be attached to the HMDs (see Fig. ??a and Fig. ??b). The diameter of the fiducial distribution was about 15-25 cm. To handle

occlusion effects between AR-based visualizations and real objects as a means to increase the depth perception, we have modeled the shelves as VRML Objects. The models of the shelves were rather coarse, only providing the polygons of the shelf that were relevant to handle occlusion effects. We used this occlusion object not only for the occlusion but also for the fine alignment of real and virtual world. For that we rendered the occlusion object in color and displayed it in the HMD, so we could visibly adjust the model with the shelf using a ruler (see Fig. ??). For the HMD optical see-through calibration we used SPAAM [?], with 20 2D/3D-correspondences and achieved an overall accuracy high enough to provide good occlusion effects in most parts of the tracked area (compare Fig ??).

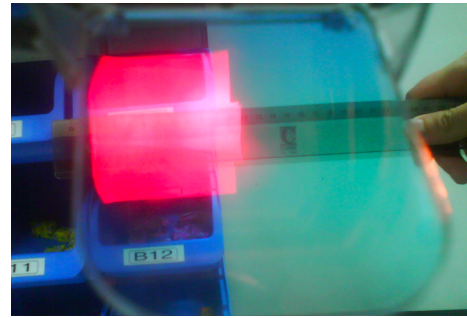


Figure 2: The colored occlusion object to execute the fine calibration, by optical aligning the objects. The ruler indicates: The virtual object has to be moved 2cm to the left. On the left side the FoV cuts the rest of the occlusion objects.

## 2 THE FIRST TEST SERIES

In our first user study we conducted an explorative experiment how to support the picking process with different displays (HMD vs. stationary monitors vs. PDA) and several visualizations (1D vs. 2D vs. 3D) on each of these displays. We here provide only a very brief overview of the 3D visualization schemes for HMDs since these formed the basis for the further investigations that are reported in this paper. Details on the remaining schemes can be found in [?, ?].

### 2.1 Experimental Setup

We used a light-weight laptop computer with a WiFi connection and mounted it in on a backpack, carried by the subjects. The display device was a monoscopic Sony Glasstron HMD (see Fig. ??). We set up two separate experiments: one for *pathfinding* (going to the correct shelf) and one for *picking* (finding the correct item on the shelf). The warehouse consisted of a shelf, 5 boxes wide and 10 boxes high for the picking experiment, and of 4 double-sided shelves with  $5 \times 5$  boxes on each side for the pathfinding experiment (see Fig. ??b). The shelves were numbered sequentially.

#### 2.1.1 Choice of Visualization

We provided visual support for three navigation tasks.

**Meta Navigation** We used a rubber band visualization to tell users where to turn their heads when the picking target was outside their field of view. The rubber band consisted of a compass-like arrow at a fixed distance of about 20 cm in front of the user's nose. The arrow pointed to the next relevant augmentation and was extended with a rubber band of flexible length (see Fig. ??b and ??b).

**Pathfinding Navigation** We set up a virtual traffic sign in front of the shelf with the next target (see Fig. ??a ). The traffic sign could be seen from everywhere, as all shelves were maximally 140 cm high.

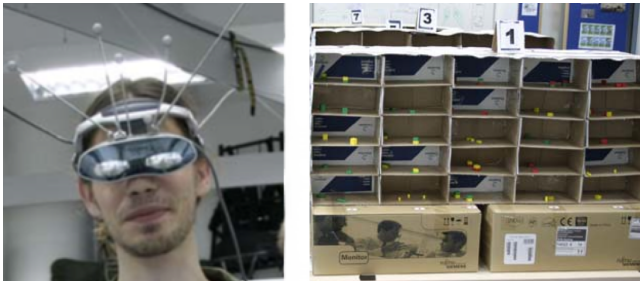


Figure 3: Setup of the first evaluation a) Sony Glasstron HMD, with retro reflective markers. b) Four double sided shelves for the pathfinding test.

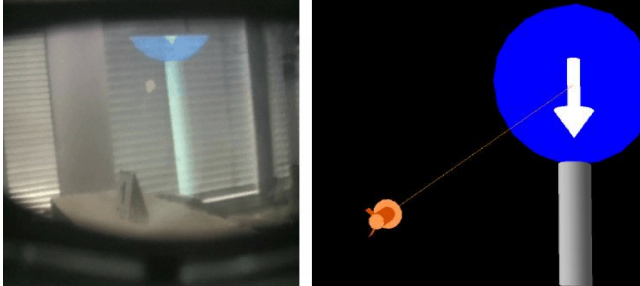


Figure 4: Pathfinding visualization a) A photograph through the Sony Glasstron. b) The rendered image

**Picking Navigation** We implemented a 3D arrow that slightly immersed into the box of the next target (Fig. ??a). The arrow was similar to the one used for the meta visualization.

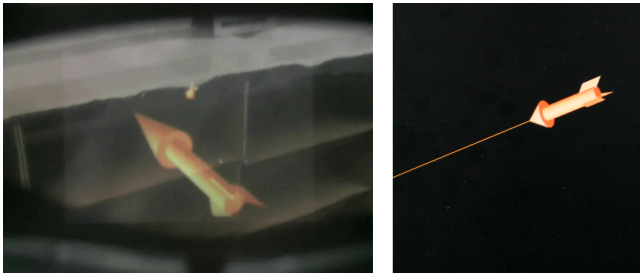


Figure 5: Picking visualization of the first test: a) Large arrow in front of picking target. b) Small arrow in front of the user's nose plus rubber band

## 2.2 Test Method

In two separate experiments for the *pathfinding* and *picking* tasks in a within-subject design, 18 subjects (age 23–48) performed each test in permuted order. We measured (dependent variables) *Way Time* + *Dead Time* for the pathfinding evaluations and *Picking Time* + *Dead Time* for picking evaluations. Additionally, we investigated the number of picking errors the participants made. The subjects were observed during the tests and had to fill out questionnaires afterward. We expected users to be faster when they were provided with 3D information than with 2D and 1D information.

**Pathfinding** The participants had to go to a specified column of the shelf and say "I have arrived" (without picking an item), always starting from an initial position in the lab. Each basic batch of tests consisted of 3 pathfinding orders listing 6 positions (i.e.

18 operations) for each visualization. For each display the same batches were used, permuted for different persons.

**Picking** The subjects had to pick a specified number of items from the correct box on a shelf and then say "I picked". Each batch of tests consisted of 3 picking orders listing 10 items. Subjects had to run a batch on the HMD for each visualization. Subjects thus were requested to make  $3 \times 10 = 30$  picks.

## 2.3 Results and First Conclusions

For pathfinding, we could not show any significant differences in time between tests with 1D or 2D visualizations, while the 3D visualization was several seconds slower than the 1D and 2D visualizations. We observed two interesting effects: First, the subjects were almost 50% faster in the third cycle than in the first. This was caused by a learning effect and by an increasing fascination of novel AR users. Second, there were two distinct groups of fast and slow users. The sub-clusters correlated with subjects who had already prior exposure to computer games or AR/VR applications and those who did not. There were only a few pathfinding failures, mostly when test persons had misunderstood the initial instructions.

In the picking test, we could not measure significant speed differences between the 1D, 2D and 3D visualizations. However, failure rates showed clear differences. Subjects made up to 10 times more mistakes with the AR-based 3D visualizations in the HMD than with the 1D/2D visualization. In most cases, subjects picked items one row too high or too low, indicating a problem with depth perception using AR. We showed AR-based picking later at a small fair using larger shelves (boxes  $40\text{cm} \times 45\text{cm}$  (*width*  $\times$  *height*) rather than  $25\text{cm} \times 15\text{cm}$ ). Here, people made almost no mistakes. That means the AR-based system had a lower bound for the box size. Below this limit, box identification became ambiguous. Comments in the subjective questionnaires showed that this is partially due to the visualization scheme and partially due to the insufficient optical see-through quality of our HMD.

Furthermore, we learned that people need some time to become comfortable with Augmented Reality. After having overcome this obstacle, people told, that they found our metaphors intuitive. People who were unexperienced with new technologies such as augmented/virtual reality, needed even more time to use the AR-based system efficiently.

## 3 THE SECOND TEST SERIES

We discussed the results of the first test series with several logistics experts/workers and came to two conclusions: 1) from the economic point of view, it is not worth supporting the pathfinding process with expensive technologies such as Augmented Reality and 2) it is highly interesting to support the picking process with AR, but it needs to be improved to prevent workers from making mistakes. We thus focused in the following on improvements to the picking visualization. To this end we developed new ways of indicating the box from which items had to be picked in 3D. We now report from the resulting new experiment. As an additional challenge, we decreased the size of the boxes on the shelves.

### 3.1 Experimental Setup

We thought about improving the depth perception by using a stereo display. But the only stereo display we had available were the i-glasses from virtual IO, with even worse see-through capabilities than the Sony Glasstron display. However we had a monocular Nomad (see Fig. ??a) display available with very high see-through capabilities, as it uses a laser projection into the retina instead of LCD technology. Additionally the Nomad does not hide the peripheral vision as the Sony Glasstron does, since it just uses a simple glass plate as an optical combiner.

Furthermore we switched to a new shelf (Fig ??b) 12 boxes wide and 8 high resulting in 96 boxes to pick from. The boxes have a



size of about  $10\text{cm} \times 10\text{cm}$  and are thus smaller than in the first experiment.



Figure 6: Setup of since the second evaluation a) Nomad from Microvision - just uses a small glass plate in the Field-Of-View b) The new shelves which are much smaller than in the first test.

### 3.1.1 Choice of Visualization

In this experiment we developed and compared three metaphors: an arrow, a rectangular frame and a tunnel.

**The Arrow** For reference, we used the arrow visualization from the first test, together with the rubber band meta visualization (see Fig. ??).

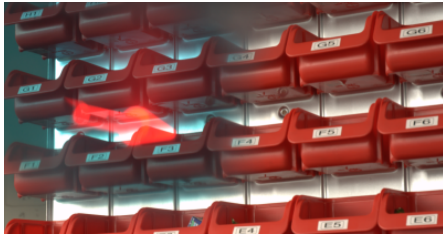


Figure 7: The arrow in front of a box (photographed through the Nomad HMD)

**The Frame** Several test participants recommended using a "simple" highlighting of the box by a rectangular frame. The implementation of this frame can be seen in Fig. ?. We again attached the rubber band.



Figure 8: The frame highlighting a box (photographed through the Nomad HMD). The frame is partly occluded by a virtual occlusion object of the shelf.

**The Tunnel** The use of a tunnel metaphor is a well known 3D concept for guiding pilots using head-up displays in airplanes. It does not matter whether the elements of the visualization symbolizing the tunnel are connected or not [?]. Biocca et al [?] adapted this metaphor to Augmented Reality-based picking tasks. They used the tunnel to visibly link the head-centered coordinate space directly to an object-centered coordinate space. They aligned the tunnel by

aligning elements on a Bezier curve between the HMD and the object to pick. In a user study they showed, that a tunnel is much faster than just highlighting an object in 3D. Thus, they basically found an alternative to our rubber band metaphor for directing the user's attention while using an HMD with a small field of view. However, in their use case they did not need the exact navigation, as our boxes are much closer together than the objects in their experiment.

In both scenarios (flight navigation and picking) the tunnel was found to be a preferred solution for rapidly guiding the user's gaze in  $4\pi$  steradians. Furthermore, it has minimal attention demands and minimizes the mental workload.

To this end, we implemented the tunnel to indicate the box to pick from. As this visualization includes the meta navigation, we did not use the rubber-band in this case. Our implementation is shown in Fig. ??.



Figure 9: The Tunnel visualization (photographed through the Nomad HMD). The blue shadow originates from the optical combiner of the Nomad.

## 3.2 Hypothesis

For the second round of experimentation we set up formal hypotheses. We developed the three different visualizations to find a better solution for the picking support. As we did not have an idea in advance, whether the new visualizations would be better or worse than the Tunnel or the Frame, we just set up two undirected hypotheses, that there will be a difference between the visualizations according to time and failures:  $H_1$  Error and  $H_2$  Time.

- $H_1$  Error:  $\exists : \mu_i \neq \mu'_i$ , with number of Errors for:  $\mu_A = \text{Arrow}$ ,  $\mu_F = \text{Frame}$ ,  $\mu_T = \text{Tunnel}$
- $H_2$  Time:  $\exists : \mu_i \neq \mu'_i$ , with Times for:  $\mu_A = \text{Arrow}$ ,  $\mu_F = \text{Frame}$ ,  $\mu_T = \text{Tunnel}$

Which leads to the Null-Hypotheses that there will not be a difference:

- $H_0$  Error:  $\mu_A = \mu_F = \mu_T$
- $H_0$  Time:  $\mu_A = \mu_F = \mu_T$

In the analysis of the experiment, we had to reject the Null-Hypotheses  $H_0$  Error and  $H_0$  Time to accept our hypotheses  $H_1$  Error and  $H_2$  Time. This rejection was evaluated for the  $\alpha$  niveau of 5%, which is a common assumption for this kind of evaluation scenarios.[?].

## 3.3 Test Method

We designed an experiment comparable to the first one. The single independent variable was the visualization. The three levels of our variable were the Arrow, the Frame and the Tunnel. We used 34 subjects between 15 and 49 years (mean age: 27.3, StdDev: 5.8, 24 men, 10 women) in a within-subject design. Half of them were acquired at the campus the other half were people from the city.



Each of them received some candies (the picked items) as reimbursement. This test was part of a comprehensive study. For this part each subject spent about 30 minutes.

We only tested the picking task and not the pathfinding. People were asked to stand in front of the shelf and to pick items. It was not possible for them to see the entire shelf while staying in front of it, so people had to move their head to see the boxes to pick from. As the subjects did not need to move a lot, we did the test without making the people wear a backpack. Instead we connected the HMD by wire to the server. Each subject got an introduction with three items for each visualization to play around with and to be able to ask questions about anything they did not understand fully. We did this intensely to compensate for the effect of fascination, which gave us a high variability in the measured results in the first experiment. The subjects had to perform 3 orders with 9 items for each visualization. For all three visualizations we used the same orders, which was not obvious to the subjects. The order in which the subjects had to use the visualizations was permuted to compensate for learning effects.

The subjects had to start each order turned backwards to the shelf. Upon a start signal they turned around and executed the test. When they said "Picked It!" we switched to the next visualization by pressing a button (Wizard of Oz technique). The system logged the time at this point and a simple harmonious sound from the speakers indicated the change in visualization. We had learned from other experiments that the latter is quite important in order not to confuse the subjects, when they do not press the button by themselves.

### 3.4 Results

For our dependent variables time and failure, we could not see (while applying ANOVA<sup>2</sup>) any significant difference between the three orders people executed. This is true for all three visualizations. Fig. ?? shows this result exemplarily for the times. The chart for the errors was not worth to be drawn. We think this positive effect (of not having huge learning/ confusing effects) between the different orders can be linked back to the introduction (try and ask) phases we gave to our subjects each time after having changed the visualization. On this account we can treat each order equally in the following analysis. In the following we discuss the measurements of the dependent variables Errors and Picking time and discuss further observations.

#### 3.4.1 Errors

We had 33 valid measurement series for the Arrow and the Frame and 34 for the Tunnel. The mean error subjects made per pick amounted to 0.165 (StDev: 0.307) for the Arrow, 0.0 (StDev: 0.0) for the Frame and 0.058 (StDev: 0.042) for the Tunnel. The results can be seen in Fig. ?. We applied a single-factor ANOVA as we just had one independent variable. It reveals a significant difference  $p = 0.0 < 0.05 = \alpha$  between the errors. This rejects our Hypothesis  $H_0$ Error, that there is no difference between the visualizations with respect to the errors. To see where the difference is, we did a post-hoc analysis. As the Levene test for the homogeneity of variances shows us, that they are homogeneous, we can apply a LSD-test (Least Significant Difference), to get further details about the differences. This gave us the following results. Subjects using the Arrow fail significantly  $p = 0.0 < 0.05 = \alpha$  more often than those using the Frame, according to mean errors 0.165. The same is true for subjects using the Arrow compared to subjects using the Tunnel, which is significant with  $p = 0.001 < 0.05 = \alpha$ , and a mean difference of 0.156 wrong items per picked item. The test rejected the difference between Tunnel and Frame with  $p = 0.822 > 0.05 = \alpha$ .

<sup>2</sup>All statistical analyses were calculated with the SPSS statistics software: [www.spss.com](http://www.spss.com)

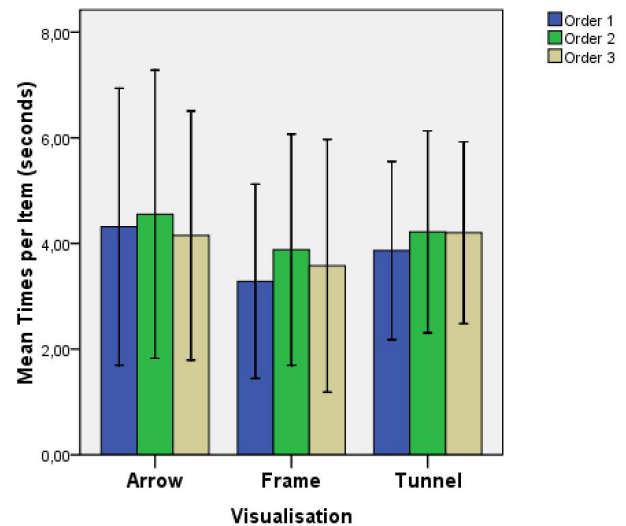


Figure 10: The measured times do not vary significantly between the orders 1-3.

It is important to mention that this does not mean there is no difference. We can just say, that it is too small to be detected in our experiment. Larger experiments with many more subjects are needed, if further clarifications are required.

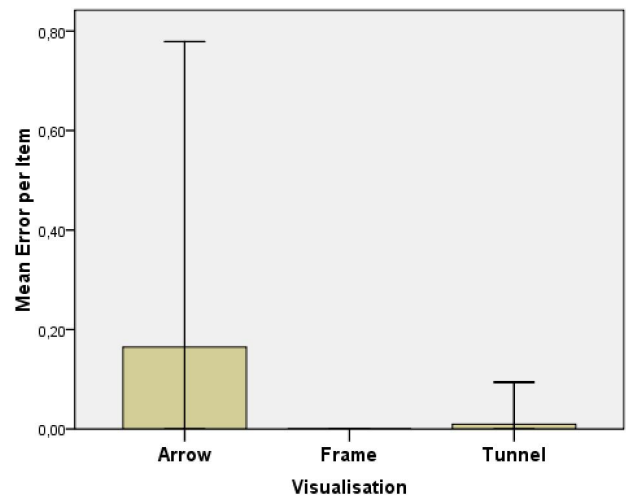


Figure 11: The mean error for each visualization, measured over all three batches. The error bar displays the double of the StDev. Frame and Tunnel differ significantly from the Arrow.

#### 3.4.2 Picking Times

We had 33 valid measurement series for the Arrow, 32 for the Frame and 34 for the Tunnel. From this we calculated the following picking times per item: 4.341s (StdDev: 1.211) for the Arrow, 3.581s (StdDev: 1.023) for the Frame and 4.096s (StdDev: 0.834) for the tunnel. The results can be found in Fig. ?. Again we applied a single-factor ANOVA and found a significant difference, with  $p = 0.013 < 0.05 = \alpha$ . This rejects our hypothesis  $H_0$ Time, that there is no difference between the visualizations with respect to the time people need, when dealing with it. To chose

the right method for the post-hoc analysis, we again applied the Levene test for the homogeneity of variances. This time the test concludes in inhomogeneity ( $p = 0.120 > 0.05 = \alpha$ ). Therefore we can not apply the LSD-test, but rather the Tamhane test, which gives us more details about the differences of the visualizations. Subjects using the Frame are significantly faster than using the Arrow. This is on average 0.76s (StdErr: 0.278) per picked item, with  $p = 0.024 < 0.05 = \alpha$ . The test rejected the differences between Arrow and Tunnel with  $p = 0.713 > 0.05 = \alpha$  and between Tunnel and Frame with  $p = 0.085 > 0.05 = \alpha$ . Again this just means that the possible difference was too small to be measured with this experiment.

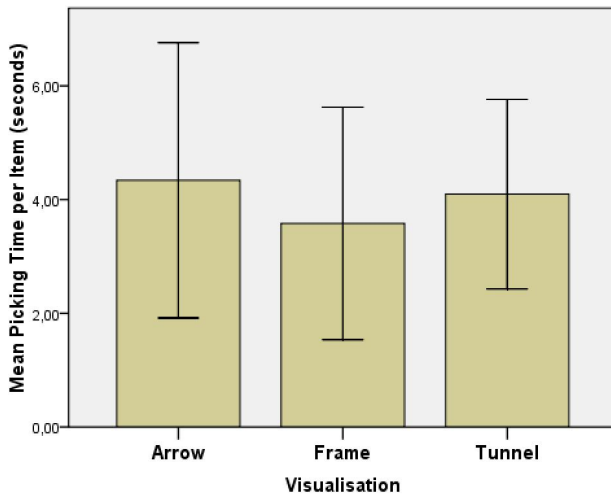


Figure 12: The mean time to pick an item for each visualization, measured over all three batches. The error bar displays the double of the StdDev. The time for the Frame differs significantly from the Arrow.

### 3.4.3 Subjective Observations

In addition to the objective measurements we now report on the observations we made during the experiment. According to Bortz[?] is it important not just to reject or accept hypotheses only on the basis of tests regarding statistical significance.

We were able to separate the subjects exposed to the Arrow visualization into two groups. Subjects of the one group moved their head around the augmented arrow until they were sure about the correct box. They even got down on their knees to be on the same height with the arrow. While (time consumingly) examining the arrow from the top, the back and the side one can figure out the right position. Thus in most cases, they picked out of the right box. However the latter is not the intention of an intuitive Augmented Reality visualization. The other group of subjects just immediately picked without great examinations of the position of the arrow. Naturally such subjects were faster, but made more mistakes.

Additionally we observed that some people closed the eye, which was not covered by the HMD. Those subjects felt uncomfortable doing this and complained about cluttering of their vision. After telling them to open both eyes they felt much better and were able to perform the test.

### 3.5 Further Conclusions

In addition to proving that the subjects were significantly slower using the Arrow than using the Frame, we determined that they made more mistakes using the Arrow than with the Frame or the Tunnel. This measurement is even more important as we do not have the benefit of a fast visualization if it is not bulletproof.

We used the same arrow visualization but with even smaller shelves than we used in the first visualization and the subjects made fewer mistakes with it. It is not possible to draw a conclusion regarding the speed because we used different shelves in both experiments. We achieved this better result for the picking process just by switching from the Sony to the Nomad HMD. We are not sure yet, where the exact reason is. It could be either because of the better see-through capabilities or because the new display does not hide the rest of the field of view and provides much more peripheral vision than the Sony does. However the participants still made mistakes using the arrow. The tunnel worked much better, but still people made mistakes, whereas the frame worked without mistakes.

In interviews many people argued, that they preferred the tunnel over the rubber band. The rubber band indicates the direction but gives not an indication how far the object is away and how fast they can turn. The Tunnel instead intuitively shows them by the strength of the bending of the Bezier curve, how far and fast they have to turn. Furthermore it immediately gives feedback about getting closer or further away. Yet, the subjects mentioned the Tunnel to be helpful for the coarse navigation but not for the fine navigation as it was sometimes ambiguous on which box it pointed. For the fine navigation they preferred the frame, because it was more exact. This relates directly to the fact, that people did not make mistakes using the frame.

From these results we combined the advantages of the visualizations by combining the frame and the tunnel (and removing the rubber band). This visualization can be seen in Fig. ??.

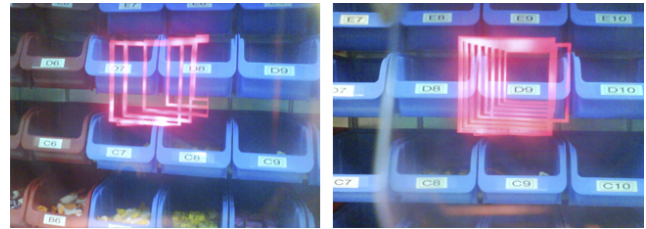


Figure 13: Two perspectives of the opaque Tunnel with squares in combination with the Frame. One sees the cluttering, the tunnel blinds the frame.

## 4 NEXT PUBLIC EVALUATION AT AN EXHIBITION

We implemented the thus concluded combined visualization. This gave us a good meta visualization to guide the user's view to the actual augmentation by using the tunnel. Furthermore this visualization is exact as we use the frame for highlighting the box. Altogether this visualization is supposed to be the fastest as we combined the two fastest visualizations of our test.

We presented this system again at some exhibitions. After a very short instruction almost all people were able to easily handle our visualization. We just explained: "There is a tunnel starting in front of your face and ending at the box from which you then have to pick the candy." People did not make mistakes but some of them complained about being disturbed by the tunnel when they wanted to finally pick the item. We observed again that some people were moving their heads around the box to see if they were seeing the visualization in front of the correct box - an effect we already observed for some people using the Arrow. However this time, the people, who did not move their head around the box, did not pick wrongly. Thus, some people had problems identifying the Frame behind the Tunnel. The problem: we simply cluttered the visualization too much.

## 5 THE THIRD TEST SERIES

With the experiences of our previous experiments and some new ideas in mind, we designed a new experiment. The main intention of this experiment was to speed up the process by reducing the cluttering. We tried to achieve this by dynamically adjusting the meta visualization, depending on the presence of the actual picking target in the users' field of view.

### 5.1 Experimental Setup

The experimental setup consisted of the components which we already used in the second test series (section ??): The Nomad-HMD and the shelves consisting of 96 boxes. The only difference was to attach the Powermate from Griffin Technology as a game show like buzzer. This device was placed at a fixed position in front of the shelf and replaced our previous Wizard of Oz-based method to step through the picking list.

#### 5.1.1 Choice of Visualization

In the last experiment and in the exhibition people never picked from a wrong box, when the box was highlighted by a frame (compare Fig. ??), whereas the simple arrow and the Tunnel alone were no safe indicators for the right box (compare Fig. ?? and Fig. ??). Due to this fact, all new visualizations based on the Frame and we concentrated mainly on optimizing the meta visualizations.

**S-Tunnel (opaque) with Frame** As first navigation we chose the square-based tunnel in combination with the frame (Fig. ??) as it was presented by us on the public exhibition (section ??). This visualization metaphor was part of our evaluation even if we already knew about the drawbacks of cluttering. We wanted to see how it performed compared to the other visualizations.

**S-Tunnel (semi-transparent) with Frame** To compensate for the disadvantage of occluding the Frame by the opaque square-based Tunnel we slightly modified the latter visualization: The purpose of the tunnel was to bring the next relevant augmentation to the field of view. That means guiding the user's gaze until he had found the Frame augmentation in front of the box. After the user had found the Frame he did not need the Tunnel anymore. So we simply faded it out when the Frame was in the user's field of view. As Biocca et al[?] proposed, the fading was done according to the dot product between the vector of the user's view and the vector pointing out of the box. This visualization can be seen in Fig. ??.

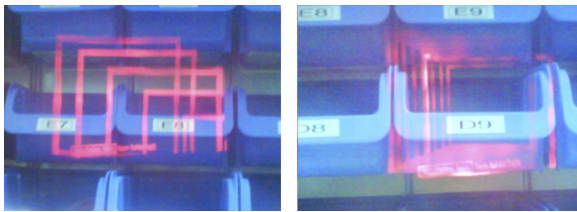


Figure 14: The semi-square tunnel. The tunnel fades out, when the real visualization is in the FieldOfView.

**Arrow with Frame** As we found the Frame in combination with the meta visualization (consisting of Arrow and rubber band) in the second test series to be good, we used it in this evaluation as a measure of ground truth.

**R-Tunnel (semi-transparent) with Frame** As we had some problems with artifacts and the overlay of several semi-transparent square corners when fading the tunnel out, we replaced the squares of the tunnel by rings. So we designed a tunnel of rings in combination with the Frame (see Fig. ??a+b). The rings were fading to transparent on base of the same function we already used for the square based tunnel.

## 5.2 Hypotheses

The intention of this experiment was similar to the last experiment. We were looking for the fastest visualization, which furthermore supported an error free picking. As the visualizations using the Frame from the last experiment, showed to be error proof and in this experiment all our visualizations used the Frame, we set up the audacious hypothesis, that people would not make an error using our visualization (with  $\mu_i$  = number of Errors for:  $\mu_A$  = S-Tunnel (opaque),  $\mu_B$  = S-Tunnel (semi-transparent),  $\mu_C$  = Arrow w. Frame,  $\mu_D$  = R-Tunnel(semi-transparent)).

- $H_1$  Error:  $\forall : \mu_i = 0$

Additionally we set up the undirected hypothesis, that there would be a difference between the visualizations according to the time (with  $\mu_i$  =Times for:  $\mu_A$  = S-Tunnel (opaque),  $\mu_B$  = S-Tunnel (semi-transparent),  $\mu_C$  = Arrow w. Frame,  $\mu_D$  = R-Tunnel(opaque)).

- $H_2$  Time:  $\exists : \mu_i \neq \mu_i'$

We could have used the experiences of the previous evaluations to set up more specific hypotheses, for example that the opaque Tunnel would be slower than the semi-transparent Tunnel. The probability to prove a directed hypothesis is much higher than proving an undirected one. We abandoned this idea, as from our point of view, this would have looked like cheating with statistics. However the Null-Hypothesis, which needs to be rejected to prove.<sup>3</sup> Our hypothesis are:

- $H_0$  Error:  $\mu_A > 0, \mu_B > 0, \mu_C > 0, \mu_D > 0,$
- $H_0$  Time:  $\mu_A = \mu_B = \mu_C = \mu_D$

## 5.3 Test Method

We basically used the same test setup as in the second experiment. The single independent variable was again the visualization. The four levels of our variable were the S-Tunnel (opaque) wF, S-Tunnel (semi-transparent) wF, Arrow wF and the R-Tunnel (semi-transparent) wF. We used 14 subjects between 20 and 50 years (mean age: 27,7, StdDev: 7,5) in a within-subject design. Each of them got again a candy as reimbursement. This time we decided to take less subjects, but spend more time observing and interviewing the people. Thus each subject spent about 45 minutes on the test.

We again only tested the picking task. Before starting with each visualization there was an introduction phase consisting of three orders for the subjects to pick (try-and-ask-phase). This introduction is, as we learned from the earlier experiment, a good way to compensate for the confusing/ learning effect resulting in long picking times in the first order. We used the same orders as in the second experiment: Each subject had to execute 3 orders with 9 items for each of the 4 visualization. The fact that the three orders were the same for all four visualizations was again not realized by the subjects. The order in which the subjects had to use the visualizations was permuted to compensate for learning effects.

The subjects this time controlled the work flow by themselves by pressing a button, placed in front of the shelf. Again we provided auditory feedback of the state change, by playing a sound when the button was pressed. As dependent variables we logged the times when people pressed the button and recorded the number of failures people made

<sup>3</sup>The rejection is evaluated for the  $\alpha$  niveau of 5%.



## 5.4 Results

We conducted some pre-evaluations of the data to check for leaning effects between the three orders for each visualization. Applying an ANOVA we could not find any significant difference for the dependent variables (time, error) between the three orders. The measured times are shown in Fig. ?? . For the errors, it was not worth drawing a chart. That means we again did not have a significant learning effect between the three orders, which allows us to treat all orders the same.

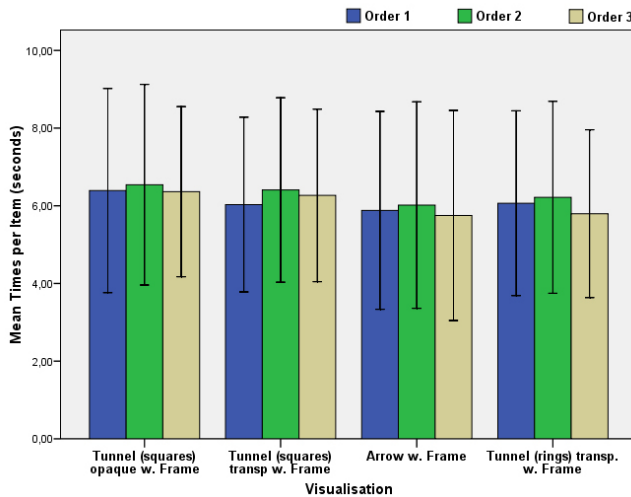


Figure 15: For all four types of visualization the measured times do not vary significantly between the orders 1-3.

### 5.4.1 Errors

14 of 14 measurement series for each type of visualization were valid. During all the 1512 picks none of the subjects picked an item wrongly, 10 times the button was pressed twice. Accordingly, we reject our Hypothesis  $H_0$  Error and, in consequence, we can accept the  $H_1$  Error, stating that people would not make any errors with one of our visualizations. We hypothesized this due to the fact that they did not make one with the Frame in the second test. However we were still surprised, that people really did not make a mistake using any meta visualization in combination with the Frame.

### 5.4.2 Picking Times

As for the errors, we had 14 of 14 valid measurement series for each visualization, leaving out the ten times when the button was pressed twice. Thus we could calculate the following picking times per item: 6,602s (StdDev:1.529s) for the S-Tunnel (opaque) wF, 6,265s (StdDev:1.213s) for the S-Tunnel (semi-transparent) wF, 6,0382s (StdDev:2.33s) Arrow wF and 6,039s (StdDev:1.202s) for the R-Tunnel (semi-transparent) wF. The results can be found in Fig. ?? . There is a difference of about two seconds in the average picking times of the second experiment compared to this one. As the only thing we changed between the two experiment was the user input (Wizard of Oz to manual switch by our subjects) this must be the cause.

However we could show significant differences between the measured picking times in this experiment, with  $p = 0.000 < 0.05 = \alpha$ . This rejects our  $H_0$ Time and accepts  $H_1$ Time that there is a difference between the visualizations in respect to the time people need to pick items. For the post-hoc analysis we again applied the Levene test for the homogeneity of variances, showing that they are homogeneous ( $p = 0.003 < 0.05 = \alpha$ ) and we for that reason can apply an LSD-test, to obtain further details about the differences

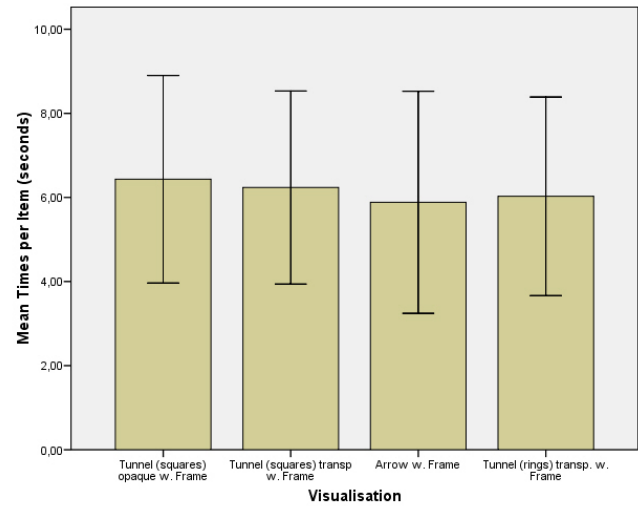


Figure 16: The mean time to pick an item for each visualization to pick an item, measured over all three batches. The error bar displays the double of the StdDev. The opaque S-Tunnel w.F. is significantly slower than the other visualizations.

between the visualizations: People performed significantly slower using the opaque S-Tunnel w.F: semi-transparent R-Tunnel on average 0.34s (StdErr:0,12s) for the S-Tunnel (semi-transparent), with  $p = 0.005 < 0.05 = \alpha$ , on average 0.56s (StdErr:0,12s) for the Arrow wF, with  $p = 0.000 < 0.05 = \alpha$  and on average 0.56s (StdErr:0,12s) for the R-Tunnel (semi-transparent) wF, with  $p = 0.000 < 0.05 = \alpha$ . A significant difference between the S-Tunnel (semi-transparent) wF and the Arrow and R-Tunnel could not be shown, it was rejected just about by  $p = 0.06 > 0.05 = \alpha$ . So it seems there is a difference, but the effect is too small to show it with the number of test persons. There was definitely no difference between the Arrow wF and the R-Tunnel (semi-transparent) wF ( $p = 0.989 > 0.05 = \alpha$ ).

### 5.4.3 Subjective Observations

In this section we will discuss the subjective observations we made during the experiment and got in the subsequent interview. Most of these facts coincide with the just discussed objective results, or are at least not contrary.

The subjects complained about the opaque S-Tunnel with Frame as it cluttered the view and they had to look around it. For that reason it was ranged as the worst solution. This fact is directly linked to the significant speed difference compared to the three other visualizations. The semi-transparent S-Tunnel with Frame was ranged as second worst visualization.

As subjectively best solution the semi-transparent R-Tunnel with Frame and the Arrow with Frame were chosen. Which again correlates with the objectively measured facts. A clear best solution was indeterminable due to several arguments. Some people argued that they preferred the Arrow over the R-Tunnel, due to less cluttering. At first sight, this is a confusing result as the tunnel is faded semi-transparent and should not clutter the Frame. However we used the proposed solution of Biocca et al [?] to fade the tunnel according to the dot-product of the start and end vector of the Bezier curve. This works fine if one looks straight at the box, but if one looks for example from above at a box which is quite low, the Frame is in the center of the HMD even though the transparency function does not work at the moment. The problem is visualised in Fig. ??a). This goes along with the fact, that subjects liked the R-Tunnel more than the S-Tunnel, because they could distinguish more easily the Rings

from the Frame. We got several interesting remarks regarding the difference of Tunnel and Arrow. Subjects liked the arrow, because it directly indicated where to look. However it did not provide any distance information. So people sometimes moved their heads too fast and looked past the Frame and then had to move the head back. Whereas the Tunnel is no good indicator for the direction at the beginning of the movement, it later provides better information about the remaining distance to the target. Furthermore people argued that sometimes they had to take one step back to see more elements of the tunnel, so the tunnel was no good indicator if one stayed directly (about 0.5m) in front of the shelf. One subject considered the tunnel to be more ergonomic, while one using the arrow was forced to make robot like movements.

## 5.5 Conclusion

In the third iteration of user studies we could prove the method of displaying the frame in front of the box as a clear technique to indicate the box to pick from.

Furthermore we could improve the meta visualization, the one which indicates where to look and go, if the actual visualization (the Frame) is not currently in the HMD's field of view. The idea to fade the Tunnel to semi-transparent when the target was in the field of view was proved to be good, but the way we did it was not. Fading on the basis of the dot product between start and end tangent of the Bezier curve does not work in some cases. On account of this, there is need for improvement. First we thought about somehow calculating the presence of the Frame in the field of view and then combining it with the current fading function. But we rejected this in favor of the following idea: We dynamically orient the vector which protrudes perpendicularly from the box, toward the user. This solves the fading problem when looking at the box with a skew. The problem and its solution are visualized in Fig. ??.

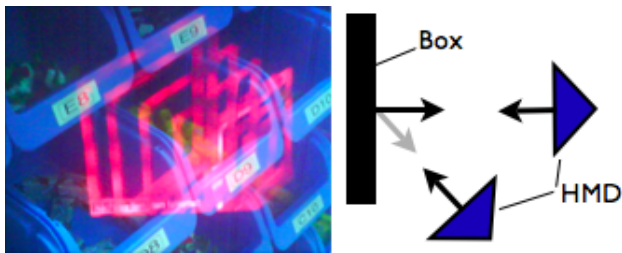


Figure 17: a) Shows the bad behaviour of the fading function when looking from a skew. b) The figure shows the problem of our first fading function. From the straight perspective the fading (dot product between both vectors) works. From a skewed viewing position the fading does not work, even if the object is close to the line of sight. The Tunnel produces clutter. Dynamically directing the vector protruding from the box toward the user solves this problem - the gray arrow.

The clear winners of this evaluation are the opaque R-Tunnel wF and the Arrow wF. However the R-Tunnel already performed (using the bad fading function) as well as the Arrow, due to the fact that the subjects could distinguish between the elements of the Tunnel and the Frame because of the geometrical difference and not the effect of transparency. The Arrow was good for giving direction information in particular when staying directly in front of the shelf. The Tunnel, on the other hand, gave a good feedback about the distance. Nonetheless it was more valuable when staying a step back from the shelf.

As a side effect, while executing this experiment we could again see the benefit of our try-and-ask phase at the beginning of each experiment. Applying this technique, we could compensate for the learning and fascination effects while executing several orders in a row.

## 6 FINAL CONCLUSION AND FUTURE WORK

In an iterative process of exploring, evaluating and refining, we found an intuitive and clear visualization, to support the order picking process. We achieved this via the execution of three empirical user studies, accompanied by presenting the system on several exhibitions. The feedback we got by observing the people at the exhibitions was at least as valuable as the feedback from the user studies.

We started displaying an Arrow to highlight boxes to pick from. Subjects had serious problems to locate the Arrow in 3D, even though we attached fins at the end of the arrow to provide a better depth perception. They still picked the items from the wrong box. We improved the highlighting of the box by displaying a Frame instead of the Arrow using the Nomad HMD. We tested this Frame in two experiments in combination with different meta visualizations and people made not a single mistake. So finally, we succeeded in designing a reliable system to precisely highlight boxes for the order picking process. Something we would not have thought after the plenty of errors people made in our first experiment, when we displayed an arrow in the Sony Glasstron HMD.

After finding the right way to highlight the box, the meta navigation is still questionable. We basically compared an arrow in combination with a rubber band versus a Bezier curve-based Tunnel to guide the users' attention to off-screen visualizations. We figured out that the Tunnel had to behave in an attentive way in order to not clutter the display: it had to disappear (fading to transparent), when it is currently not needed. The first fading function did not behave in an ideal way, so we proposed a new optimized one. Due to the bad fading function we concluded that the R-Tunnel was better than the S-Tunnel as one could distinguish the rings better from the squares. In any case the transparent R-Tunnel wF performed as fast as the Arrow wF, even if the fading function was bad. We could show that both visualizations are well suited for our small warehouse, where people do not have to move a lot - with the Arrow having its strengths in the directional and the Tunnel in the distance information. Currently we are setting up an experiment in a larger warehouse in which people move much more with large distances between the shelves, and user change in a wide spectrum. We hope this will lead us to more conclusions about the Arrow and the Tunnel. Furthermore, we are evaluating a combination of Arrow and Tunnel (without the rubber band). Beside this we will focus in future evaluations on the mental workload. We did not do this in the past, since we first needed to know which visualization works well, before handing out further questionnaires.

As a by-product of our iterative experiments we could successfully test our try-and-ask introduction before each test. Each participant got an introduction with three items to pick in advance. He was allowed to try the new visualization out, until he understood it. In the first experiment, we did not apply this technique and we got extreme learning and fascination effects from people trying Augmented Reality for the first time, resulting in time differences of 50% from the first to the last order. Applying this technique fully compensated for these effects in the second and third experiment.

## ACKNOWLEDGEMENTS

The authors wish to thank T. Frimor, E. Yükselgil, M. Stadter, X. Pan, R. Reif and our test persons, who executed all those thousands of picks for us (and ate all the candy). This work was supported within the ForLog-Project by Abayfor.