Evaluating the Benefits of Real-time Feedback in Mobile Augmented Reality with Hand-held Devices

Can Liu[1,2,3] **Stéphane Huot**[2,3] **Jonathan Diehl**[1] **Wendy E. Mackay**[3,2] **Michel Beaudouin-Lafon**[2,3]
{can.liu , jonathan.diehl}@rwth-aachen.de {huot , mackay , mbl}@lri.fr

[1] RWTH Aachen University
52056 Aachen, Germany

[2] Univ Paris-Sud & CNRS (LRI)
F-91405 Orsay, France

[3] INRIA
F-91405 Orsay, France

ABSTRACT
Augmented Reality (AR) has been proved useful to guide operational tasks in professional domains by reducing the shift of attention between instructions and physical objects. Modern smartphones make it possible to use such techniques in everyday tasks, but raise new challenges for the usability of AR in this context: small screen, occlusion, operation "through a lens". We address these problems by adding real-time feedback to the AR overlay. We conducted a controlled experiment comparing AR with and without feedback, and with standard textual and graphical instructions. Results show significant benefits for mobile AR with feedback and reveals some problems with the other techniques.

Author Keywords
Augmented Reality; Hand-held Devices; Real-time Feedback.

ACM Classification Keywords
H.5.1 Information Interfaces and Presentation: Multimedia Information Systems—*Artificial, augmented, and virtual realities*

INTRODUCTION
Augmented Reality (AR) techniques have been used to assist workers in complex tasks such as assembly [7] or maintenance [3] by overlaying localized instructions onto physical objects with a Head Mounted Display (HMD). But casual users may also need guidance when using common appliances: *"How do I set this washing machine for this cloth?"* While the use of a HMD is impractical for daily activities, recent smartphones can be used for "Mobile Augmented Reality" applications and provide in-place guidance anywhere and for everyone (Fig. 1).

Manipulating objects while following instructions is an "Alternating Attention" task [6] that requires to switch between two complementary sub-tasks. In our context, notes and instructions are usually made of plain text or pictures to be interpreted before operating the device. Text instructions require an additional level of interpretation than pictures to map the descriptions to actual objects. Conversely, pictures provide a better spatial mapping to retrieve physical objects but require annotations to convey additional information, e.g., how to operate a control, which value to use for a setting. In both cases, when instructions are long or complex, repeatedly switching between sub-tasks can be highly demanding for the user as it requires memorizing instructions, visually finding the objects of interest, retrieving the next instruction, etc.

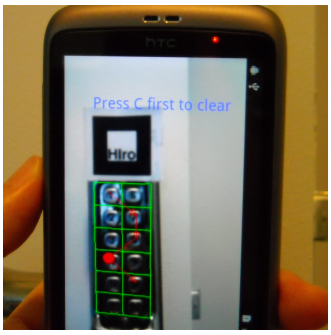


Figure 1. Hand-held AR instructions for entering a code on a keypad.

For such tasks, AR techniques combine the advantages of text and pictures as they ease the localization of physical objects by displaying additional "in-place" information [3]. This should improve user performance by reducing task switching and alternating attention. However, occlusions from the AR layer and the user's hand while interacting with physical objects may offset this benefit. This is even more likely with hand-held devices, due to their small size.

In order to take full advantage of mobile AR for instruction tasks, we propose to provide real-time feedback of actions in the physical world directly in the AR layer. Combining instructions and real-time status of physical objects on top of the camera input, e.g., the status of a button or the position of a knob, solves the occlusion problem and should enable interaction through the device screen, thus minimizing task-switching. After discussing related work, the rest of this paper presents a mobile AR prototype with feedback (AR+F) and a study that assesses the advantages of this approach.

RELATED WORK
The benefits of AR for assisting operational tasks have been evaluated in different contexts. Feiner et al. demonstrated the concept in their early KARMA system [1]. Tang et al. compared the efficiency of a HMD-based AR system in an object-assembly task with printed manual and overlaid instructions

C. Liu, S. Huot, J. Diehl, W. E. Mackay, M. Beaudouin-Lafon.
Evaluating the Benefits of Real-time Feedback in Mobile Augmented Reality with Hand-held Devices.
In CHI'12: Proceedings of the 30th International Conference on Human Factors in Computing Systems, ACM, May 2012.
Authors Version

on a LCD display [7]. They found that AR instructions reduce error rate and cognitive effort, suggesting that AR assists mental transformation and minimizes attention switching. In the context of vehicle maintenance, Henderson and Feiner showed that a HMD AR system helps to retrieve the object of interest and also assists the psychomotor phase of procedural tasks better than static 3D instructions on a monitor [3].

Using mobile devices for see-through AR in operational tasks has received less attention. Hakkarainen et al. describe a mobile AR system for assembly tasks [2] that displays sequences of image in adjusted real-world perspective. AAR [5] uses a similar approach by displaying pre-rendered animations on a location-tracked tablet-PC to learn a complex machine.

Providing real-time AR feedback of physical actions in the real world has received little attention so far. Apart from Kotranza et al's Mixed-Reality medical learning system where students receive real-time feedback about their actions [4] and Henderson and Feiner's work [3], we are not aware of other similar work in Mobile AR nor of any empirical study about the effectiveness of this approach.

In summary, most studies have focused on specialized tasks with expert or specialized users and with specific devices (HMD). While they demonstrate the potential of AR, its benefits for everyday manipulation of physical appliances by casual users with standard mobile devices, with or without real-time feedback, remain to be demonstrated.

AR WITH REAL-TIME FEEDBACK PROTOTYPE

Providing real-time feedback in mobile AR applications raises technical issues that require state-of-the-art technologies in ubiquitous computing and computer vision, whether the manipulated objects and the mobile device are connected or not. We can expect such technologies to be available on mobile phones in the near future. In the mean time, and since our primary objective is to evaluate and understand the benefits of this approach, we built an ad-hoc prototype based on a generic controller that communicates with a mobile device.

We use a JLCooper CS-10[2] MIDI control station (Fig. 2) that features controls commonly found on physical appliances (buttons, knobs, sliders and a jog wheel). The mobile application runs on a HTC Desire mobile phone running Android 2.3. We use the NyARToolkit[1] for tracking fiducial markers attached to the control panel. Data from the controller (the values of the controls) is transmitted in real-time to the mobile application through the MIDI and OSC protocols.

When the mobile phone recognizes a marker, it displays an AR layer on top of the real-time camera image with outlines of the physical controls and in-place instructions to perform the task, e.g. a value to enter or a button to press. The user can move the controls while looking through the phone. A color overlay provides real-time visual feedback: if the user is expected to set slider 3, a red bar indicates the value to dial (Fig. 2c&d). Once she moves the correct slider to the correct value, the bar turns blue (Fig. 2d). If she uses the wrong control, its outline turns purple.

[1] http://nyatla.jp/nyartoolkit/wiki/index.php

ASSESSING MOBILE AR FOR OPERATING INSTRUCTIONS

We conducted a controlled experiment to assess the benefits of hand-held mobile augmented reality instructions for manipulating physical devices, with or without real-time feedback. Participants were asked to perform tasks of variable difficulty using our prototype. In addition to *Augmented Reality* (AR) and *Augmented Reality + Feedback* (AR+F), we also tested *Text* and *Picture* instructions as control.

Based on previous studies and our experience with the prototype, our hypotheses are:

- *H1:* With respect to speed, (a) AR+F outperforms AR; (b) AR+F and AR outperform Text and Picture; (c) Performance differences increase with task difficulty.
- *H2:* AR techniques are less error-prone and facilitate the correction of errors.
- *H3:* AR+F instructions are preferred to Text, Pictures and AR instructions by users.

Participants. We recruited twelve men and four women, all right-handed, age between 24 and 44. None of them had any experience with augmented reality applications, but four were frequent users of surface controllers, e.g., mixing consoles or guitar amplifiers, and nine owned a smart phone.

Apparatus. Settings were performed on the controller of our prototype and the instructions were displayed on a HTC Desire mobile phone (display: 3.7 inches, resolution: 480×800 px, weight: 135 g, dimensions: 119×60×11.9 mm), running Android 2.3. Text and Picture instructions were displayed with standard Android widgets and AR techniques were performed with our prototype. We used 10 buttons, 6 knobs and 8 faders of the controller, which was covered with white paper to hide the other controls and display AR markers (Fig. 2).

Design and Procedure

The experiment is a [4×3] within-subject design with two factors: TECHNIQUE (*Text*, *Picture*, *AR*, *AR+F*) and DIFFICULTY (*Easy*, *Medium*, *Hard*). A trial is a "setting" task requiring participants to set the controls on the panel as instructed by the corresponding TECHNIQUE on the mobile device, in any order. Once done, participants are asked to press a hardware button on the device. In case of failure, they are asked to continue the trial and correct the setting until they succeed or a timeout occurs. At the end of the trial, the participants reset the controller and rest in a default position in front of the panel before starting the next trial.

Instructions for each Technique.

– *Text* instructions are displayed with a control-value pair on each line (Fig. 2a), as in mobile note-taking apps. To avoid a potential order effect inherent to linear presentations, the order of lines is randomized across participants.

– *Picture* instructions are presented with a 1024 × 537 pixels image of the expected setting (Fig. 2b). The picture is initially fully visible and participants can pan and zoom to get a better view of the controls.

– *AR* displays vector graphics on top of the camera input and highlights every control of the panel (Fig. 2c). Controls that must be set appear in red together with the target value.

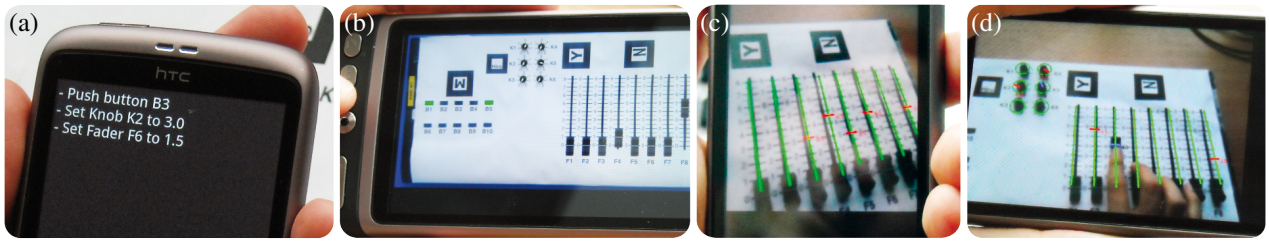


Figure 2. Setting instructions methods: a) Text. b) Picture. c) Augmented Reality (AR). d) Augmented Reality with real-time Feedback (AR+F).

– *AR+F* is similar to *AR* but updates in real-time as controls are manipulated by the user (Fig. 2d): the current value of the manipulated control is displayed next to it and the overlay turns blue once set to the correct value.

Settings and Difficulty. Each setting task includes three types of controls – buttons, knobs and sliders – and its difficulty is determined by the total number of controls to manipulate. For each type of controls, the range of possible values is (i) 0 or 1 for buttons, (ii) 1 to 5 for knobs and (iii) 1 to 7 for sliders, both with a 0.5 interval for the latter two.

To ensure that the number of controls to manipulate in a setting is the only parameter affecting difficulty, the following rules are used to randomly generate each setting: (i) each control is used at most once; (ii) the number of controls of each type is balanced; (iii) the values of continuous controls are balanced (same number of low, medium and high values).

We conducted an informal pilot study with two participants and six difficulty levels (3, 6, 9, 12, 15, 18 controls) to validate these rules and determine appropriate difficulty levels for the study. It suggested that settings with the same number of controls are equivalent in terms of performance and that three levels of difficulty are sufficient: 3 controls for *Easy*, 9 for *Medium*, 18 for *Hard*, with timeouts of 2, 4, and 6 minutes.

Finally, to avoid learning effects, we generated several random settings for each difficulty level and counterbalanced these settings across participants so that none of them perform the same setting twice during the experiment.

Procedure. Trials are grouped into blocks according to TECHNIQUE and each TECHNIQUE × DIFFICULTY condition is replicated twice. The four techniques are introduced to the participants at the beginning of the experiment. Participants perform practice trials with a low difficulty level until they feel comfortable with each technique and the control panel. Then, they perform one *recall* trial prior to each new TECHNIQUE. The presentation order for TECHNIQUE and DIFFICULTY is counterbalanced across participants using a Latin square. This design results in (4 × TECHNIQUE) × (3 × DIFFICULTY) × (2 replications) × (16 × PARTICIPANT) = 384 measured trials.

Each session lasted about 40–50 minutes, after which participants where asked to rank each technique in general and for each difficulty level.

Data Collection. We collected: (i) *TrialTime*, the trial completion time; (ii) *ReactionTime*, the time from the appearance of the instructions to the first action of the participant on the controller; (iii) *Errors*, the number of errors by trial, i.e. the number of times a wrong setting was validated during a trial.

Results and Discussion

We removed 3.03% outliers (trials with a total time greater than two standard deviations from the mean) and performed a full-factorial analysis TECHNIQUE×DIFFICULTY×Random(PARTICIPANT) with the REML technique for repeated measures and Tukey HSD for post-hoc tests.

H1: Which Technique Performs Best?

We found a significant effect of TECHNIQUE ($F_{3,45} = 45.61$, $p < 0.0001$) and DIFFICULTY ($F_{2,30} = 299.83$, $p < 0.0001$) and a significant TECHNIQUE×DIFFICULTY interaction effect ($F_{6,90} = 27.59$, $p < 0.0001$) on *TrialTime*. As expected, *TrialTime* increases with DIFFICULTY and all difficulty levels are significantly different (*Easy* = $11.8s$, *Medium* = $28.4s$, *Hard* = $56s$). For TECHNIQUE, *AR+F* ($23.9s$) is significantly faster than all other techniques, and *Text* ($43.9s$) is significantly slower (*AR*= $29.4s$ and *Picture*= $31.1s$). This result supports most of *H1a&b* as *AR* outperforms *Text* but not *Picture*.

The small difference in performance between *AR* and *Picture* could be explained by the screen resolution and the graphical occlusions by the hand that hinder manipulation while looking through the device's screen, reducing the potential benefit of displaying in-place instructions. In fact, we observed that participants were using *AR* in a similar way to *Picture*: (i) they held the AR display over the panel for an overview and moved it for a close-up look at several controls in order to memorize the values; (ii) they moved the AR display aside to interact with the controls directly; (iii) they moved it back to learn the next set of controls. By contrast, *AR+F* allows to set the controls while keeping the attention focused on the on-screen instructions thanks to its real-time feedback, thus increasing performance by reducing the division of attention.

The TECHNIQUE×DIFFICULTY interaction effect provides a more interesting insight into the performance of each technique according to task difficulty (*H1c*). As shown in Fig. 3, *TrialTime* increases sharply with difficulty for *Text*, more slowly for *AR+F*, with *AR* and *Picture* in between. This is confirmed by a TECHNIQUE×Random(PARTICIPANT) ANOVA, with significant effects of TECHNIQUE for each DIFFICULTY level:

- At the *Easy* level, only *Picture* is performing slower than other instruction methods. This is likely due to the time required to find the controls to set in the picture, compared with reading and memorizing three textual instructions or skimming through the panel with the AR techniques;
- At the *Medium* level, the performance of *Text* drops significantly and *AR+F* performs faster than *Picture*. *Text*, *Picture* and *AR* require more attention switches because instructions are more difficult to memorize (nine controls). Also, *Text* probably makes it more difficult to keep track of the instructions in the list while switching attention;

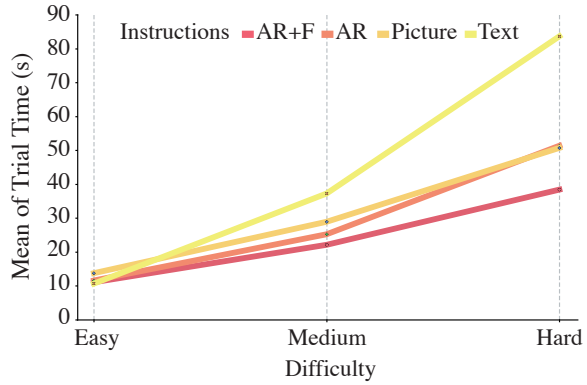


Figure 3. Mean of *TrialTime* for each TECHNIQUE by DIFFICULTY.

- At the *Hard* level, these differences are exacerbated. *Text* is by far the slowest technique, *Picture* and *AR* have similar performance, and *AR+F* is the fastest (Fig. 3).

H2: Are AR Techniques Less Error-Prone?

13.4% of trials had at least one error. A nominal logistic ANOVA model for TECHNIQUE×DIFFICULTY∼*Errors* shows significant effects for TECHNIQUE ($\chi^2 = 11.39, p = 0.0098$) and DIFFICULTY ($\chi^2 = 30.62, p < 0.0001$). We observe an increase in error rate with DIFFICULTY, however post-hoc tests are only significant for *Hard* (26.6%) against *Medium* (11.7%) and *Easy* (3.12%). For TECHNIQUE, the only significant difference is between *Picture* (21.9%) and *AR+F* (5.2%). Error rates for *Text* and *AR* are 15.6% and 12.5% respectively.

These results partially support *H2*: while *AR+F* and *AR* have fewer errors, not all differences are significant. Also, we do not find any effect of TECHNIQUE or DIFFICULTY on the number of errors per trial that were not successful on the first attempt. So we cannot conclude about error recovery. In fact, assessing error recovery would have required a specific error correction task. Finally, we observe that error rate alone cannot explain the difference in performance among techniques since it does not exhibit the same effects as *TrialTime*. This supports our assumption that alternating between subtasks is the most influential performance factor.

H3: Do Participants Prefer AR Techniques?

Participants were asked to rank the instruction methods (1 to 4) for each difficulty level and overall. A nominal logistic ANOVA model for TECHNIQUE×DIFFICULTY∼*Rating* shows a significant effect of TECHNIQUE ($\chi^2 = 233.61, p < 0.0001$) and a significant TECHNIQUE×DIFFICULTY interaction effect ($\chi^2 = 46.81, p = 0.0104$). In fact, *AR+F* was ranked as the preferred technique by all participants overall and by most of them for each difficulty level. It was followed by *AR*, *Picture* and *Text*, except at the *Easy* level where *Text* was often ranked 3rd, explaining the interaction effect.

Participants preferred *AR+F* mainly because it does not require to switch between the device and the panel: the interactive feedback solves the occlusion problem with *AR*. In fact, ten users never looked at the panel with *AR+F* and four only occasionally, while only two looked at it all the time. For *AR*, participants mostly raised the issue of graphical occlusion, but also explained that without interactive feedback, the size of the screen and the resolution of the camera image impaired precise manipulation when not looking at the panel. Some participants also felt unnatural to set the controls through the screen with both AR techniques, but this was balanced by the benefits of real-time feedback for *AR+F*.

Overall, these ratings and comments are consistent with our quantitative analysis. Participants' preferences match their performance, and their comments support the same explanations that led to our hypotheses.

CONCLUSION AND FUTURE WORK

We presented a prototype and a user study to evaluate the performance of AR on mobile phones for daily operational tasks. We simulated a technique – AR+Feedback – that adds real-time feedback to the AR overlay and compared it with Text, Picture and AR without real-time feedback for task instructions. The results show that the conventional AR approach performs almost as well as pictures, and that AR+Feedback significantly improves task performance and user experience. More studies are needed to explore real-time AR feedback for hand-free operations with a HMD, or in comparison with other kinds of dynamic instructions, e.g., Text and Pictures with real-time feedback. However the present work provides valuable insights to incorporate feedback to mobile AR systems.

This work leads to a new way for AR to go out of professional domains and reach a larger audience. We will apply and test this concept with more use cases and other types of tasks. We also plan to develop further prototypes to improve the connection between mobile devices and physical objects.

ACKNOWLEDGEMENTS

We thank Clément Pillias, Pierre Rossinès, Theophanis Tsandilas and David Bonnet for their help and suggestions, the anonymous reviewers for their feedback, and our participants for their time and efforts.

REFERENCES

1. Feiner, S., Macintyre, B., and Seligmann, D. Knowledge-based augmented reality. *CACM 36* (1993), 53–62.
2. Hakkarainen, M., Billinghurst, M., and Woodward, C. Augmented assembly using a mobile phone. In *MUM '08*, ACM (2008), 84–87.
3. Henderson, S. J., and Feiner, S. K. Augmented reality in the psychomotor phase of a procedural task. In *ISMAR '11*, IEEE (2011), 191–200.
4. Kotranza, A., Scott Lind, D., Pugh, C., and Lok, B. Real-time in-situ visual feedback of task performance in mixed environments for learning joint psychomotor-cognitive tasks. In *ISMAR '09*, IEEE (2009), 125 –134.
5. Quarles, J., Lampotang, S., Fischler, I., Fishwick, P., and Lok, B. A mixed reality approach for merging abstract and concrete knowledge. In *VR '08*, IEEE (2008), 27 –34.
6. Sohlberg, M. M., and Mateer, C. A. *Introduction to Cognitive Rehabilitation: Theory and Practice.* Guilford Press, 1989.
7. Tang, A., Owen, C., Biocca, F., and Mou, W. Comparative effectiveness of augmented reality in object assembly. In *CHI '03*, ACM (2003), 73–80.