

## Research Article

# WozARd: A Wizard of Oz Method for Wearable Augmented Reality Interaction—A Pilot Study

Günter Alce,<sup>1</sup> Mattias Wallergård,<sup>2</sup> and Klas Hermodsson<sup>1</sup>

<sup>1</sup>Sony Mobile Communications, Nya Vattentornet, 221 88 Lund, Sweden

<sup>2</sup>Lund University, P.O. Box 118, 221 00 Lund, Sweden

Correspondence should be addressed to Günter Alce; [gunter.alce@sonymobile.com](mailto:gunter.alce@sonymobile.com)

Received 28 November 2014; Revised 22 April 2015; Accepted 26 May 2015

Academic Editor: Francesco Bellotti

Copyright © 2015 Günter Alce et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Head-mounted displays and other wearable devices open up for innovative types of interaction for wearable augmented reality (AR). However, to design and evaluate these new types of AR user interfaces, it is essential to quickly simulate undeveloped components of the system and collect feedback from potential users early in the design process. One way of doing this is the wizard of Oz (WOZ) method. The basic idea behind WOZ is to create the illusion of a working system by having a human operator, performing some or all of the system's functions. WozARd is a WOZ method developed for wearable AR interaction. The presented pilot study was an initial investigation of the capability of the WozARd method to simulate an AR city tour. Qualitative and quantitative data were collected from 21 participants performing a simulated AR city tour. The data analysis focused on seven categories that can have an impact on how the WozARd method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, general user-experience, and human-operator performance. Overall, the results indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human-operator mistakes.

## 1. Introduction

The age of wearable devices is upon us and they are available in many different form factors including head-mounted displays (HMDs), smartwatches, and smartbands [1]. Wearable devices are intended to always be “on,” always acting, and always sensing the surrounding environment to offer a better interface to the real world [2]. Taking into account recent advances in wearable devices, we can expect that people will be able to carry their wearables at all times. One example of a wearable form factor that follows this trend is HMDs. HMDs have been developed and used in research since the 1960s [3], but it is not until recently that they have become available outside of the research lab. Examples of HMDs or glasses that are available are *Google Glass* [4], *Meta-Pro* [5], *Recon Jet* [6], *Vuzix M100* [7], and *Epson Moverio BT-200* [8].

The HMD form factor facilitates augmented reality (AR), a technology that mixes virtual content with the users' view of the world around them [9]. Azuma [10] defines AR as having three characteristics: (1) combining real and virtual, (2) being

interactive in real time, and (3) being registered in 3D. According to Narzt et al. [11] the AR paradigm opens innovative interaction facilities to users: human natural familiarity with the physical environment and physical objects defines the basic principles for exchanging data between the virtual and the real worlds, thus allowing gestures, body language, movement, gaze, and physical awareness to trigger events in the AR space. However, it is difficult and time consuming to prototype and evaluate this new design space due to components that are undeveloped or not sufficiently advanced [12]. To overcome this dilemma and focus on the design and evaluation of new user interfaces (UIs) instead, it is essential to be able to quickly simulate undeveloped components of the system in order to enable the collection of valuable feedback from potential users early on in the design process. One way of doing this is with the Wizard of Oz (WOZ) method. The basic idea behind WOZ is to create the illusion of a working system. The person (puppet) using it is unaware that some or all of the system's functions are actually being performed by a human operator (wizard), hidden somewhere “behind

the screen.” This allows testing interaction concepts before a system is fully working. The method was initially developed by Kelley in 1983 to simulate a natural language application [13].

The WOZ method has been used in a variety of studies to explore design concepts for interactive systems. An early application area was simulating speech recognition systems [14].

Another application area in which it is suitable to use the WOZ method includes AR [15]. A WOZ tool called *DART* [16] enables designers to design AR UIs and to integrate live video, tracking technology, and other sensor data. Lee and Billinghurst [17] used WOZ to study multimodal AR interfaces but only in an indoor static setup.

Although WOZ has been used for a long time and in various application areas, there is still no WOZ tool known by the authors that can be used to prototype AR UIs that work in both indoor and outdoor environments and that can be used with HMDs and other wearable devices integrated with a mobile phone (e.g., based on Android) for mobility.

The authors have developed a WOZ tool called *WozARd* in an attempt to meet these requirements. The set of features that *WozARd* offers is described in more detail in [18]. With *WozARd* it is possible to control what is shown and played on the user’s HMD and/or smartphone or tablet. *WozARd* lets the user interact with the system through a smartwatch. The human operator can easily change the UI without reprogramming the application, which makes *WozARd* flexible and easy to use for nonprogrammers.

One important aspect when using the WOZ method is to ascertain that the participants’ behavior in the simulated system is reasonably similar to that in the corresponding real system [14]. The extent to which a study comprises “real-world” use of a system is called “ecological validity” [19]. Another term which is closely related to ecological validity is “external validity,” which means the extent to which the results of a study can be generalized to other situations [20].

For example, low fidelity prototyping such as paper prototyping has a low ecological validity, but it can be very effective in testing issues of aesthetics and standard graphical UI. In other words, by using low fidelity prototyping with low ecological validity, it is still possible to achieve high external validity. However, to do so when designing for an ecosystem of wearable devices, a richer ecological validity is often required [19].

The context in which *WozARd* was developed was the three-year European project VENTURI [21]. The project’s first year focused on AR gaming, the second year was about supporting visually impaired people, and the third year had AR city tours as theme. The goal of the third year was to deliver an AR application that let people experience a city’s cultural heritage, through their own smartphones and/or tablets. Part of the objective was also to allow participants to experience parts of the city tour with HMD. For this reason *WozARd* was developed and used within VENTURI to explore fundamental design issues connected to AR city tours early on in the project.

Furthermore, AR navigation systems such as “The Touring Machine” developed by Feiner et al. [22], Narzt et al. [11],

and Bolter et al. [23] research were used as inspiration for this study.

The goal of the presented pilot study was to perform an initial investigation of the capability of the *WozARd* method to simulate a believable illusion of a real working AR city tour. Mainly aspects concerning the method itself were studied, but also the limitations of current hardware were considered since they contribute to the participants’ experience.

The study presented was carried out by collecting and analyzing qualitative and quantitative data from 21 participants who performed a predefined city tour using *WozARd* on wearable devices. The data analysis focused on six categories that are believed to have a potential impact on how the *WozARd* method is perceived by participants: precision, relevance, responsiveness, technical stability, visual fidelity, and general user experience.

The next section presents relevant related work. Then the *WozARd* tool is described followed by a presentation of the method, results, discussion, conclusions, and future work.

## 2. Related Work

As mentioned, WOZ is a well-known method where a human operates undeveloped components of a technical system. Above all, the WOZ method has been widely used in the field of human-computer interaction to explore design concepts. WOZ testing is a powerful method for uncovering design ideas in limited evolved components, especially for systems performing in physical environments, since the designers are less constrained by technical specifications [24]. Dow et al. [25] state that WOZ testing is “leading to more frequent testing of design ideas and, hopefully, to better end-user experiences.”

An early application area of WOZ was in speech recognition systems [14]. To simulate both input and output language technology components, Schlögl et al. developed an open-source tool called *WebWOZ* [26] that uses an internet based WOZ framework.

The WOZ method has also been used to combine speech and gestures to control a robot by speech and gestural interaction [27]. Two human wizards were used in the evaluation, one responsible for the dialogue and the other for the robot navigation. Other gesture based WOZ studies include [28].

Other examples of research tools that used the WOZ method include *ConWIZ* [29], which is a WOZ tool with a mobile application that is capable of controlling the simulation of a WOZ prototype as well as contextual objects such as fans and lights. Fleury et al. [30] used a WOZ setup to evaluate four different methods for transferring video content from a smartphone to a TV screen. Li et al. [24] developed *Topiary* which lets users interact with the UI mockup while a human wizard follows them and updates the locations.

As already shown, there are several WOZ tools available for different use cases. However, none of them fulfill the requirements of being flexible, mobile, able to add other form factors, and able to explore AR interaction. Some of the listed WOZ tools are flexible but not mobile [26]. Examples of mobile WOZ tools include *ConWIZ* [29], Linnell et al.’s tool [31], and *Topiary* [24], but they do not support integration of



FIGURE 1: The WozARd architectural setup.

other form factors nor can they be used for exploring AR. In most of the studies, the human operator's role is stationed in a control room hidden from the participants. However, with mobile tools such as *ConWIZ* [29], Linnell et al.'s tool [31], and *Topiary* [24], the wizard was able to follow the participants and update the UI accordingly, but it is not clear whether the participants knew that the system was controlled by the human operator or not.

None of the mentioned WOZ tools, however, fulfilled the requirements which were needed to perform studies for the VENTURI project. Examples of requirements for the VENTURI project included the following:

- (i) Do not focus on one form factor.
- (ii) Be useable both indoors and outdoors.
- (iii) Aid the human operator when adding scenarios on the fly.
- (iv) Support the easy adding of other form factors.
- (v) Be suitable for prototyping AR.

In past research by the authors, the WOZ tool WozARd [18] was developed in an attempt to meet these requirements.

### 3. The WozARd Tool

This section introduces the WozARd WOZ tool; a more detailed description of the tool can be found in [18]. First, an overview is presented on how the tool works, followed by examples of features that the tool supports.

WozARd consists of two Android devices that communicate with each other wirelessly (Figure 1).

On the left is the wizard device which is controlled by the human operator and on the right within the dashed lines are the devices used by the participant (Figure 1). Through the WozARd wizard application, the human operator can control the participant's UI by pressing the buttons in the application. Examples of features that WozARd is suitable for are

- (i) presentation of media such as images, video, and sound (Figure 2(a));
- (ii) navigation and location based triggering (Figure 2(b));
- (iii) showing notifications (Figure 2(c));
- (iv) features to plan and prepare for user studies;
- (v) capability to log test and visual feedback;
- (vi) being able to work with both tablet and phone form factors;
- (vii) integrating the Sony *Smartwatch* [32] and the Sony *Smartwatch 2* [33] for interaction possibilities;
- (viii) adding HMDs, which can be connected through HDMI, for example, *Vuzix Star 1200* [34];
- (ix) adding HMDs, which run on Android, for example, *Epson Moverio BT-200* [8], *Vuzix M100* [7], and *Google Glass* [4].

The only type of interaction that the participant can perform is touch gestures on a Sony *Smartwatch*, which catches the gesture performed by the participant and sends it through the Bluetooth connection to the wizard device. Of course, other interaction types based on, for example, voice and midair hand gestures, could be simulated as long as the human operator can hear and see the participant properly and interpret his/her intentions correctly. Figure 3 shows what a participant sees through a video see-through display when the human operator pushes the turn right button.

### 4. Method

This section describes the setup of the pilot study.

The approach to this pilot study was to first define categories that can have a potential impact on how the WozARd method was perceived by participants. The ISO definition of usability [35], which includes effectiveness, efficiency, and satisfaction, was used as starting point. Each of the three usability categories was subcategorized resulting in a total of six categories (Figure 4):

- (i) *Precision*. Is the augmented information shown at the right time and place?
- (ii) *Relevance*. Is relevant information shown at the right time and place?
- (iii) *Responsiveness*. How quickly does the system respond to user input?
- (iv) *Technical Stability*. Did the user notice any technical difficulties?
- (v) *Visual Fidelity*. What fidelity does the visual input have? Since the WozARd does not currently support tracking, it is not possible to impose virtual content correctly registered in the 3D space. Instead, the image is "hanging" in front of the user (i.e., when the user turns his/her head, the image follows the head movement).

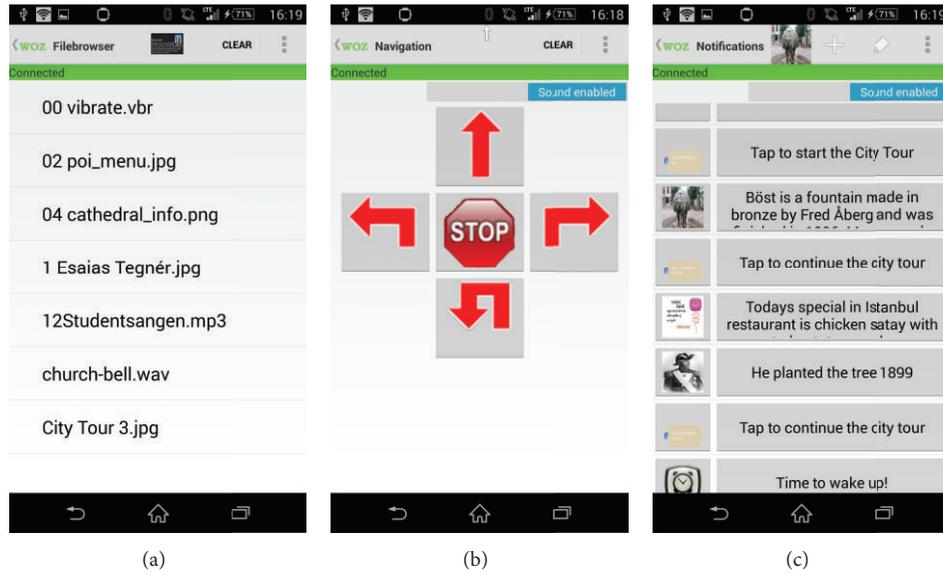


FIGURE 2: The wizard device UI: (a) list of playable multimedia, (b) navigation, and (c) notifications.

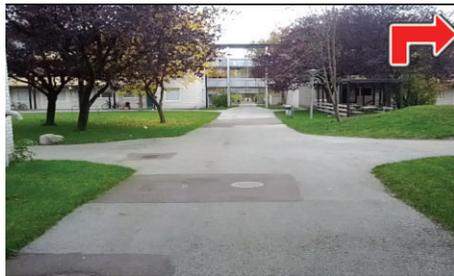


FIGURE 3: The participant’s view through a video see-through display, when the human operator pushes the turn right button.

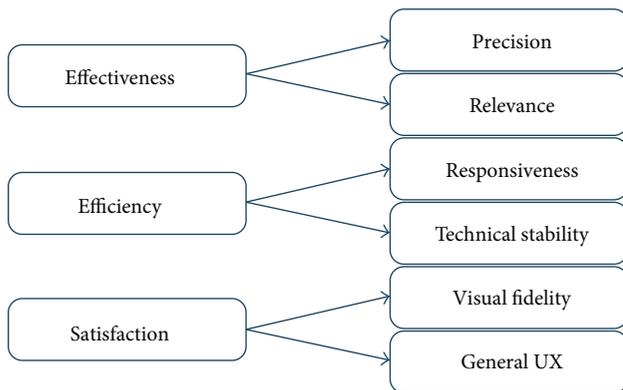


FIGURE 4: Six categories emerged from the ISO 9241-210.

(vi) *General User Experience*. What is the general user experience of the WozARd method including the ability to hear and read the augmented information?

Nine pilot experiments were conducted iteratively, which resulted in continuous improvements of the tool and the experimental setup.

The AR city tour took place in a small city in southern Sweden called Trelleborg. The tour was based on a predefined route. All information and images were collected prior to the study and included different types of urban environments and target objects. The information that the participants experienced contained an image and audio, mainly text to speech. Examples of participant experiences included historical information, informative notifications (Figure 5(a)), lunch specials at restaurants, tourist attractions (Figure 5(b)), and sculptures (Figure 5(c)). Participants had to interact with a Sony Smartwatch [32] to start the city tour, to continue the city tour, and to remove notifications.

The tour was designed to let the participants walk approximately 500 m (Figure 6(a)). The average time to walk the city tour was eight minutes.

4.1. *Materials*. Equipment used during the pilot study included

- (i) HMD, Vuzix Star 1200 [34] connected to Sony Xperia S [36];
- (ii) Sony Xperia S [36] used as puppet device;
- (iii) Sony Xperia Z [37] used as wizard device;
- (iv) Sony Smartwatch [32] used by the participants to interact with the system;
- (v) Sony Handycam HDR-CX190 [38] to record during the user study.

4.2. *Participants*. 21 participants (6 women and 15 men), mainly students, were recruited for the study. The average age was 26.2 years (SD = 14.2). The participants reported that they used computers or tablets 3.67 hours per day (SD = 2.92) and smartphone on average 4.98 hours per day (SD = 3.61).

4.3. *Procedure*. The sessions involved a participant; the human operator who simulated the AR city tour with the



FIGURE 5: Three samples of what was shown during the city tour.



FIGURE 6: (a) A map of the route, (b) the human operator on the left guiding one of the participants, and (c) a participant interacting with the system.

WozARd wizard device and managed the experiment; and a test assistant who walked along the participant and video-recorded each session for data capture and measuring elapsed time (Figure 7). The session started with the participant signing an informed consent form and filling out a background questionnaire. The questionnaire included participant age, gender, and occupation. Next, a short introduction of AR was given by describing Azuma’s definition [10], followed by instructions on how to interact with the system. The participants were also asked to follow the instructions from the system and to think aloud while walking the city tour. Thinking aloud is one of the most direct and widely used methods to gain information about participants’ internal states [39]. Using the thinking aloud method had two purposes: to gain information on the participants’ experience when attending to the information and to aid the

human operator during the city tour in understanding if the participants were experiencing any problems. In addition, participants were informed that the human operator would walk behind them taking notes.

All participants filled in a questionnaire after the tour. It contained fifteen statements inspired by the System Usability Scale (SUS) [40] to which the participant agreed or disagreed on a five-point Likert scale. The questionnaire was designed to target the six categories: precision, relevance, responsiveness, technical stability, visual fidelity, and general user experience (Table 1). Each session lasted about 30 min.

The session was concluded with an informal, open interview to collect qualitative data. Each session was video-recorded. Each participant’s video recording was transcribed with individual quotes categorized and labeled. Furthermore, events of special interest were noted, for example, human

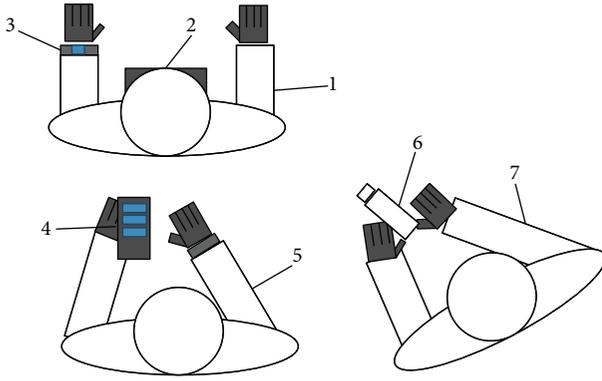


FIGURE 7: A conceptual setup of the study. 1: participant, 2: Vuzix Star 1200 [34], 3: Sony Smartwatch [32], 4: Sony Xperia Z [37] WozARd wizard side, 5: human operator, 6: Sony Handycam HDR-CX190 [38], and 7: test assistant.

TABLE 1: Questionnaire statements.

Number	Statement
S1	The system was responsive.
S2	The system was precise considering showing right picture on the right place.
S3	The city tour was interrupted due to technical difficulties.
S4	The system showed me relevant information at the right time and place.
S5	The “hanging” images were annoying.
S6	The system showed too much information too often.
S7	I enjoyed walking the city tour.
S8	The pace of the tour was appropriate.
S9	The interaction with the watch was intuitive.
S10	I would recommend the city tour to a friend.
S11	I would like to use this system when I visit a new city.
S12	It was hard to read the notifications.
S13	It was hard to hear what the system said.
S14	It felt like the system was working completely automatically.
S15	It felt like someone was controlling the system.

operator induced errors. The answers from the five-point Likert scale questionnaire responses were given as numerical value from 1 to 5 (disagree = 1; agree = 5) for the statistical calculations of median (Mdn) and interquartile range (IQR), which is the distance between the 75th and 25th percentiles.

## 5. Results

This section presents quantitative and qualitative data from the user pilot study. Overall, all of the 21 participants managed to accomplish the AR city tour and the majority of them showed signs of enjoying the AR experience. The data in the following is divided into the seven categories: precision, relevance, responsiveness, technical stability, visual fidelity, general user experience, and human operator performance.

TABLE 2: Median and min. and max. values.

Questionnaire statement	Median	Min.	Max.
S1: The system was responsive.	5	3	5
S2: The system was precise considering showing right picture on the right place.	5	3	5
S3: The city tour was interrupted due to technical difficulties.	1	1	5
S4: The system showed me relevant information on the right time and place.	5	4	5
S5: The “hanging” images were annoying.	2	1	4
S6: The system showed too much information too often.	1	1	3
S7: I enjoyed walking the city tour.	5	3	5
S8: The pace of the tour is appropriate.	5	3	5
S9: The interaction with the watch was intuitive.	5	2	5
S10: I would recommend the city tour to a friend.	5	2	5
S11: I would like to use this system when I visit a new city.	5	1	5
S12: It was hard to read the notifications.	3	1	5
S13: It was hard to hear what the system said.	1	1	4
S14: It felt like the system was working completely automatically.	5	2	5
S15: It felt like someone was controlling the system.	1	1	4

The last category was not part of the original six categories that were hypothesized to have a potential impact on how the WozARd method is perceived by participants but emerged as a new category during the data analysis.

Since the distribution was not symmetric and an ordinal scale was used, the median was calculated for the questionnaire responses [41]. The whiskers show the range of the data set, that is, max./min. value. The values of the statements are presented in Figure 8 and Table 2.

**5.1. Precision.** The majority of the participants thought that the system was precise; three participants neither agreed nor disagreed. The median value was 5 (IQR = 1).

Several participants made comments regarding precision. One participant, for example, seemed impressed that “the system knew that I was close to the street and told me to stop and then continue when I had crossed the street.” Another participant liked that the “daily specials from the restaurants popped up about 15 m before so you had a chance to think if you wanted to eat there or not.”

**5.2. Relevance.** All participants found the information they received to be relevant considering both the right time and place, that is, showing today’s offer when passing the restaurant and not what movie is shown in the cinema. This category had a median value of 5 (IQR = 0).

Positive feedback was given about the relevance of information during the interviews. One participant particularly

TABLE 3: Listing problems that occurred during the city tour.

Type of error	Description	Number of occurrences
Lost Wi-Fi connection	Notification was not sent due to loss of Wi-Fi connection. For one participant this resulted in a missed turn. The human operator, however, managed to redirect the participant. In the case of the second participant, the city tour had to be ended prematurely. The only thing missing, however, was the showing of an image and information that the tour was over.	2
Low battery	Low battery notification resulted in that the participant needed to hurry up to finish the city tour.	1
Smartwatch touch functionality	The touch functionality of the smartwatch did not register that the participant had pressed or made a swipe gesture. However, none of the eight participants noticed the lack of touch functionality since the human operator noticed the problem and sent the notification as if the touch functionality had worked.	15
Human operator error	The human operator missed sending a notification. However, the two participants showed no signs of noticing that something was missing.	2

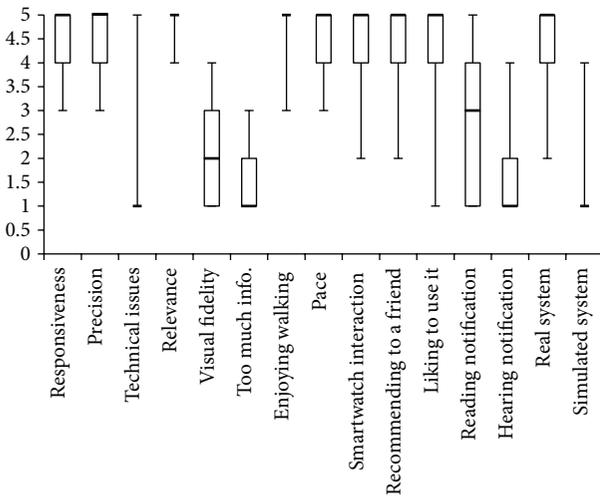


FIGURE 8: Median and the whiskers show the range of the data set.

liked that the daily menu was shown in such a way that you knew what was being offered without the need to take out the mobile phone to search for that information. Two participants, though, thought that all the information would be intrusive when they went down a street that only consisted of restaurants.

**5.3. Responsiveness.** The majority of the participants agreed that the system felt responsive. For example, one participant commented that “the system reacted immediately when I turned.” Only one participant selected neutral (i.e., neither agreed nor disagreed). The median value was 5 (IQR = 1).

**5.4. Technical Stability.** There were technical problems which the human operator needed to manage, but from the participants’ point of view there were two participants who commented on experiencing technical problems, one which resulted in aborting at the end of the city tour and for one who received low battery notification (Table 3).

**5.5. Visual Fidelity.** The input from the participants on visual fidelity was diverse, but in general they did not express extreme opinions. The mean value was 2 (IQR = 2).

The feedback was quite varied in the interviews. One participant pointed out that when the image is “hanging” in front of you (i.e., when you turn your head, the image follows the head movement) it disfigures the view of what you actually want to see. Additionally, one participant disliked the current solution and suggested that if the image was correctly placed in the 3D space, it could be used as a means of interaction (i.e., if you did not look at the building the image would disappear). Another participant, though, liked it since it helped to find the “target” that one might have missed if the augmented information was correctly registered in the 3D space.

**5.6. General User Experience.** Several statements were used to collect data about the general user experience such as S6, S7, S8, S9, S10, S11, S12, and S13 (Table 1).

The participants seemed to enjoy walking the tour; only one participant was neither positive nor negative towards the AR city tour simulated with WozARd. The amount of information was also considered to be well balanced.

The participants used the smartwatch to interact during the city tour and most of the participants found it discrete and intuitive. Four participants mentioned that they would like to be able to use speech as well.

Several participants commented on the industrial design of the glasses during the interviews. Example included the following: “it would be embarrassing, people would think that I had lost my mind,” “the design should be woman-friendly,” and “I would like to use the system when I visit a new city, if it looked nicer.”

The answers were more diverse about both hearing and reading the notifications. The main problem with reading the notifications was glare due to the sun. Ten participants reported that they had problems with the sun. One stated “I had to find a place in the shade to be able to read the notifications.”

Because some areas were crowded, some participants could not clearly hear the instructions. One user could not hear the information being presented when an ambulance passed by and also stated that it was impossible to repeat the information.

One participant reported that it was disturbing that the arrows used for navigation had 90° angles, which resulted in several unnecessary turns.

**5.7. Human Operator Performance.** During the nine pilot trials, the importance of letting the human operator see what was displayed in the participant's view was noticed. In the experimental setup it was therefore arranged so that the human operator got visual feedback from the participant side. As for audio, however, there was no feedback indicating that the audio information was being played on the participant side or when it had finished. Due to this, the human operator could accidentally interrupt the audio information by sending a new command to the participant. However, after the pilot trials, the human operator knew when the sound was about to finish and could adjust the timing of the notifications. Consequently, none of the participants reported that they had any problems with the audio information being cut short.

Despite the attempt to make the wizard device UI as usable as possible, at two occasions the human operator missed to send a notification to the participant's view (Table 3) when the two participants passed a point of interest of the AR city tour. Since the participants were unknowing about the information that should have been shown to them, they of course did not notice any problem.

Since the touch functionality of the smartwatch did not work properly and sometimes failed to catch the participant's swipe gesture, the human operator had to be proactive and react when this occurred. The human operator managed to notice and address the problem every time it occurred (Table 3).

Another concern that made it difficult for the human operator was the Wi-Fi connection, which occasionally started to fluctuate in crowded areas. However, the connection managed to stabilize quickly enough so that all information and notifications were sent to the user with one exception (Table 3).

Also power usage constituted a problem. Since the screen, Wi-Fi, and Bluetooth were always on, the phone needed to be recharged often. One misjudgment by the human operator resulted in a "low battery" notification for one participant (Table 3).

## 6. Discussion

Overall, the results seem to indicate that the participants perceived the simulated AR city tour as a relatively realistic experience despite a certain degree of technical instability and human operator mistakes. Their subjective experience of the simulated AR city tour, as measured by the questionnaire, was overall positive and in general the city tour seemed to induce a feeling of a real, autonomous system rather than a system being controlled by someone else. The observation data seemed to confirm this. All participants managed to

accomplish the AR city tour and in general they seemed to enjoy walking the simulated AR experience.

Based on the experiences of this study, the authors believe that two of the most important factors contributing to these results are the design of the wizard device of the WozARd tool and the skill of the human operator. The wizard device of WozARd was designed to aid the human operator in controlling the notifications during the user study and to reduce the risk of human operator mistakes. However, despite this, notifications to be sent to the participant's AR view were missed. This risk for mistakes could be decreased by letting WozARd's wizard device provide the human operator with visual hints that aid him/her when activating commands in the GUI (Figure 2). For example, in the notification view, already shown notifications could be grayed out and the upcoming notification in the list could be highlighted in some way.

Since the participants were walking in an outdoor environment, unpredicted turns took place and therefore the human operator needed to react accordingly. Another example of a small detail that could have easily been missed by a novice human operator was that the smartwatch [32] did not work properly and sometimes failed to catch the participant's swipe gesture. However, participants appeared not to pay attention to the problem since the human operator noticed and reacted when the touch functionality did not work. This indicates that it is important to have a skilled human operator who can control both WozARd and the test situation simultaneously in order to react to unpredictable and unexpected events. It has been suggested that the skill of the human operator is often a general problem with the WOZ method since it relies on the human operator not making any mistakes [42].

Another aspect that potentially can have a large impact on how a test participant perceives a WOZ test is the actual hardware used. The HMD [34] used in this study is one of the earliest HMDs available for early adapters. The insufficient display technology was reflected in the results concerning readability, one of the aspects that had the most diverse responses. The main reason for the participants' troubles to read the notifications was glare due to the sun. The participants who tested the tour in the afternoon reported the most problems with glare. This demonstrates a potential problem in using a WOZ tool like WozARd: the difficulties observed could be due to insufficient hardware rather than design issues connected to the AR user interface itself. The potential bias this can introduce in test results must be carefully considered when using a method like WozARd. The glare problem, nevertheless, can be expected to diminish with the development of newer display technology such as the one used in *Epson Moverio BT-200* [8]. Another aspect concerning hardware is how participants' behavior and performance may be affected by the actual industrial design. Several participants commented that they did not think that the HMD was very attractive. However, the present study did not target potential effects of the system's industrial design and we therefore only report that this is a possible source of bias to be aware of. Naturally, the use case as well as the chosen field setting of being in an outdoor environment has an

impact on how a test participant perceives an AR city tour. For example, problems such as glare from the sun, rain, crowded areas, and Wi-Fi connection could have been avoided if the study was conducted in a controlled indoor environment. However, since the VENTURI third-year theme was about AR city tours conducting the study outdoor meant a higher ecological validity and richer feedback for the project.

In the present study only one specific use case for wearable AR was simulated. No real claims about the general usefulness of the WozARd method in a design process can therefore be made based on the presented data. Even though only a few participants commented on the lack of realistic tracking of the augmented data in the present study, many use cases involve moving people and objects and would depend on a real AR tracking algorithm that correctly integrate the augmented data with the real world 3D space. For such use cases, the WozARd method might not be able to facilitate meaningful prototyping. In its current form, WozARd is probably best suited for testing scenarios limited to visualization of 2D data, tactile feedback, and auditive feedback. However, WozARd has been released as an open-source project [43], which makes it possible for others in the AR community to expand the features and tweak it to fit their own requirements and use cases.

## 7. Conclusions

In conclusion, the WozARd method seemed to work reasonably well at least for this specific use case. Further, two of the most important factors that contributed to the simulated AR experience and induced a feeling of a real, autonomous system are the design of the wizard device of the WozARd tool and the skill of the human operator. Last, although WozARd does not support AR tracking in its current form, this did not emerge as a critical problem for this particular use case. However, if another use case with, for example, moving people and objects is intended to be studied, AR tracking might be needed.

## 8. Future Work

We will continue to develop WozARd and investigate its usefulness as a prototyping tool for wearable AR. Much effort will be put into the design of the wizard device in order to help the human operator not making any mistakes. Examples of improvements include letting WozARd's wizard device provide the human operator with visual hints that aid him/her when activating commands in the GUI and adding visual feedback when the audio has stopped playing on the participant's side. Further, it is of great interest to study the importance of the human operator's level of expertise in using the WozARd tool. One way of doing this could be to investigate the difference in performance between a group of novice human operators, who will only get a short introduction to WozARd, and a group of expert human operators.

A feature that could add to WozARd's ability to facilitate meaningful prototyping is AR tracking. This could be especially useful for studies that involve moving people and

objects. Since WozARd has been released as an open-source project [43], this and other features could be developed together with other developers in the AR community.

The software development kit [44] for *Google Glass* was released after this study and WozARd has been updated to be able to run with *Google Glass*. It would thus be interesting to conduct another study in which the user only needs to wear *Google Glass* and a *Sony Smartwatch* to investigate what results WozARd could produce with better HMD hardware.

Finally, it would be interesting to investigate how WozARd can be used to explore AR UIs in other environments, such as the home, by simulating, for example, a smart living room in which the user can control consumer electronics with different wearable devices such as *Google Glass* and *Sony Smartwatch* or *SmartBand*.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This research was partially funded by the European 7th Framework Program, under Grant VENTURI (FP7-288238), and the VINNOVA funded Industrial Excellence Center EASE (Mobile Heights). The authors would like to extend their gratitude to Ertan Muedin who helped conduct the user studies. Special thanks are due to Eileen Deaner for language support.

## References

- [1] V. Genaro Motti and K. Caine, "Understanding the wearability of head-mounted devices from a human-centered perspective," in *Proceedings of the ACM International Symposium on Wearable Computers (ISWC '14)*, pp. 83–86, Seattle, Wash, USA, September 2014.
- [2] J. Rekimoto, Y. Ayatsuka, and K. Hayashi, "Augment-able reality: situated communication through physical and digital spaces," in *Proceedings of the 2nd International Symposium on Wearable Computers. Digest of Papers*, (Cat. No.98EX215), pp. 68–75, 1998.
- [3] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, Part I (AFIPS '68)*, pp. 757–764, ACM, 1968.
- [4] Google Glass, 2013, <https://www.google.com/glass/start/>.
- [5] Meta Pro, 2014, <https://www.spaceglasses.com/>.
- [6] Recon Jet, 2014, <http://www.reconinstruments.com/products/jet/>.
- [7] Vuzix M100, 2014, <http://www.vuzix.com/consumer/products-m100/>.
- [8] Article ID 560020, Epson Moverio BT-200 Smart Glasses, 2014, <http://www.epson.com/cgi-bin/Store/jsp/Product.do?sku=V11H560020>.
- [9] E. Barba, B. MacIntyre, and E. D. Mynatt, "Here we are! where are we? Locating mixed reality in the age of the smartphone," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 929–936, 2012.
- [10] R. T. Azuma, "A survey of augmented reality," *Presence*, vol. 6, no. 4, pp. 355–385, 1997.

- [11] W. Narzt, G. Pomberger, A. Ferscha et al., “Augmented reality navigation systems,” *Universal Access in the Information Society*, vol. 4, no. 3, pp. 177–187, 2006.
- [12] N. Davies, J. Landay, S. Hudson, and A. Schmidt, “Rapid prototyping for ubiquitous computing,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 15–17, 2005.
- [13] J. F. Kelley, “An empirical methodology for writing user-friendly natural language computer applications,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '83)*, pp. 193–196, Boston, Mass, USA, December 1983.
- [14] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, “Wizard of Oz studies—why and how,” *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258–266, 1993.
- [15] S. Dow, J. Lee, C. Oezbek, B. MacIntyre, J. D. Bolter, and M. Gandy, “Wizard of Oz interfaces for mixed reality applications,” in *Proceedings of the Conference on Human Factors in Computing Systems (CHI EA '05)*, pp. 1339–1342, April 2005.
- [16] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J. D. Bolter, and M. Gandy, “Wizard of Oz support throughout an iterative design process,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 18–26, 2005.
- [17] M. Lee and M. Billinghurst, “A Wizard of Oz study for an AR multimodal interface,” in *Proceedings of the 10th International Conference on Multimodal Interfaces (ICMI '08)*, pp. 249–256, October 2008.
- [18] G. Alce, K. Hermodsson, and M. Wallergård, “WozARd: a wizard of oz tool for mobile AR,” in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)*, pp. 600–605, August 2013.
- [19] S. Carter, J. Mankoff, S. R. Klemmer, and T. Matthews, “Exiting the cleanroom: on ecological validity and ubiquitous computing,” *Human-Computer Interaction*, vol. 23, no. 1, pp. 47–99, 2008.
- [20] J. E. McGrath, “Methodology matters: doing research in the behavioral and social sciences,” in *Readings in Human-Computer Interaction: Toward the Year 2000*, pp. 152–169, 1995.
- [21] P. Chippendale, P. Benjamin, D. Buhrig et al., “VENTURI—immersive Enhancement of User-world Interactions,” 2012.
- [22] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster, “Touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment,” in *Proceedings of the 1st International Symposium on Wearable Computers*, pp. 74–81, October 1997.
- [23] J. D. Bolter, M. Engberg, and B. MacIntyre, “Media studies, mobile augmented reality, and interaction design,” *Interactions*, vol. 20, no. 1, pp. 36–45, 2013.
- [24] Y. Li, J. I. Hong, and J. A. Landay, “Design challenges and principles for wizard of Oz testing of location-enhanced applications,” *IEEE Pervasive Computing*, vol. 6, no. 2, pp. 70–75, 2007.
- [25] S. Dow, B. MacIntyre, J. Lee, C. Oezbek, J. D. Bolter, and M. Gandy, “Wizard of Oz support throughout an iterative design process,” *IEEE Pervasive Computing*, vol. 4, no. 4, pp. 18–26, 2005.
- [26] S. Schlögl, G. Doherty, N. Karamanis, and S. Luz, “WebWOZ: a wizard of oz prototyping framework,” in *Proceedings of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '10)*, pp. 109–114, June 2010.
- [27] A. Green, H. Hüttenrauch, and K. S. Eklundh, “Applying the wizard-of-Oz framework to cooperative service discovery and configuration,” in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN '04)*, pp. 575–580, September 2004.
- [28] S. Connell, P.-Y. Kuo, L. Liu, and A. M. Piper, “A wizard-of-oz elicitation study examining child-defined gestures with a whole-body interface,” in *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*, pp. 277–280, June 2013.
- [29] T. Grill, O. Polacek, and M. Tscheligi, “ConWIZ: a tool supporting contextual wizard of Oz simulation,” in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*, December 2012.
- [30] A. Fleury, J. S. Pedersen, and L. Bo Larsen, “Evaluating user preferences for video transfer methods from a mobile device to a TV screen,” *Pervasive and Mobile Computing*, vol. 9, no. 2, pp. 228–241, 2013.
- [31] N. Linnell, R. Bareiss, and K. Pantic, “A wizard of Oz tool for android,” in *Proceedings of the 14th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '12)*, pp. 65–70, September 2012.
- [32] Sony SmartWatch, 2012, <http://www.sonymobile.com/global-en/products/accessories/smartwatch/>.
- [33] Sony SmartWatch 2, 2013, <http://www.sonymobile.com/global-en/products/accessories/smartwatch-2-wrist-strap-se20/>.
- [34] Vuzix Star 1200, 2010, <http://www.vuzix.com/UKSITE/augmented-reality/products.star1200.html>.
- [35] International Organization for Standardization (ISO), *ISO 9241-210, Ergonomics of Human-System Interaction—Part 210: Human-Centred Design for Interactive Systems*, International Organization for Standardization, London, UK, 2010.
- [36] Sony Xperia S, 2012, <http://www.sonymobile.com/global-en/products/phones/xperia-s/>.
- [37] Sony Xperia Z, 2013, <http://www.sonymobile.com/global-en/products/phones/xperia-z/>.
- [38] Sony High Definition Camcorder HDR-CX190, 2012, [http://store.sony.com/gsi/webstore/WFS/SNYNA-SNYUS-Site/en\\_US/-/USD/ViewProduct-Start?SKU=27-HDRCX190/B&CategoryDomainName=&CategoryName=](http://store.sony.com/gsi/webstore/WFS/SNYNA-SNYUS-Site/en_US/-/USD/ViewProduct-Start?SKU=27-HDRCX190/B&CategoryDomainName=&CategoryName=).
- [39] K. A. Ericsson and H. A. Simon, “Verbal reports as data,” *Psychological Review*, vol. 87, no. 3, pp. 215–251, 1980.
- [40] System Usability Scale (SUS), 2013, <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>.
- [41] P. K. Janert, *Data Analysis with Open Source Tools*, O'Reilly Media, Inc., Sebastopol, Calif, USA, 2011.
- [42] D. J. Lazar, D. J. H. Feng, and D. H. Hochheiser, *Research Methods in Human-Computer Interaction*, 2010.
- [43] WozARd source code, 2014, <https://github.com/sonyxperia-dev/WozARd>.
- [44] Google Developers, 2014, <https://developers.google.com/glass/develop/gdk/>.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

