ICIC Express Letters
Part B: Applications
Volume **13**, Number **5**, May **2022**
 ISSN 2185-2766
pp. 527–533

DESCRIPTION OF USABILITY PROBLEMS IN AUGMENTED REALITY USER INTERFACES

Gangseok Suh, Jongmoon Lim, Subin Shin and Wonil Hwang*

Department of Industrial and Information Systems Engineering
Soongsil University
369 Sangdo-Ro, Dongjak-Gu, Seoul 06978, Korea
{ suh; whdans9370; sinsubin1016 }@soongsil.ac.kr; *Corresponding author: wonil@ssu.ac.kr

Received September 2021; accepted December 2021

Abstract. *Finding and reporting usability problems properly through usability evaluation methods such as heuristic evaluation is an essential part of the UI development process, but the usability problems found can be often described abstractly, despite the request for detailed description. It is necessary to identify the factors that make the evaluator describe the abstract usability problems and consider them in the usability evaluation process. This study aims to explore which factors influence the evaluator's description types of usability problems. The experiments were conducted with a total of 46 participants, who were assigned randomly to use one of two heuristic types (general vs. specific) for heuristic evaluation and asked to interact with both of two types of augmented reality (AR) user interfaces (UI) (high level vs. low level of user interaction) and to describe usability problems of AR UI while conducting maintenance tasks for computer components. Based on the collected usability problems, the evaluator's description types of usability problems (abstract vs. concrete) were coded for the dependent variable. The mixed effects logistic regression was applied to finding significant factors among heuristic type, user interaction level, gender, language (Korean vs. English) and AR experience (high vs. low vs. none) affecting the evaluator's description types of usability problems. As a result of analysis, it was found that user interaction level and language are significant factors to affect the evaluator's description types of usability problems. Implications for practitioners were provided based on these experimental results.*
Keywords: Heuristic evaluation, Language, Gender, Experience, Evaluator characteristics, Augmented reality

1. **Introduction.** When evaluating a prototype of a user interface (UI), finding usability problems through usability evaluation methods is an essential part of the UI development process [1]. However, incorrect documentation or miscommunication of usability problems derived from usability evaluation can result in poor results compared to the effort invested in usability evaluation [2,18]. Usability problems are described by the evaluator during the usability evaluation process and delivered to the UI developer along with the severity of the problem. So, how well a UI developer can improve an existing UI by identifying and fixing usability problems depends in part on how the evaluators describe the usability problems they find. If the evaluator describes an abstract usability problem that is not specific, it will be more difficult for the UI developer to fix the usability problem and improve the UI than when the specific usability problem is delivered. In the process of usability evaluation, the difference in the discovered usability problems and their severity ratings depending on the evaluator is called the evaluator effect [3], and also how the discovered usability problems are described can be considered as a part of the evaluator's effect. For example, when using the heuristic evaluation method [16], which asks an evaluator to evaluate the UI from the point of view of users using the heuristics and to describe

DOI: 10.24507/icicelb.13.05.527

the usability problems in detail, in some cases the usability problems found can be described abstractly, despite the request for detailed description. Therefore, it is necessary to identify the factors that make the evaluator describe the abstract usability problems and consider them in the usability evaluation process. The purpose of this study is to explore which factors influence the evaluator's description types of usability problems. The evaluator's description types of usability problems include abstract and concrete descriptions. We conducted experiments in which evaluators were asked to find and describe usability problems of augmented reality (AR) UIs through heuristic evaluation. Based on the description of usability problems, we investigated which factors among heuristic type, user interaction level and evaluator's characteristics, such as gender, language and AR UI experience, affect the evaluator's description types (abstract description vs. concrete description). Based on the experimental results, it can be concluded that two factors, i.e., user interaction level and gender, are significant, and finally we can suggest that these factors should be taken into account to obtain a concrete description of the usability problem in the heuristic evaluation. This section introduces the motivation and purpose of this study, and is followed by a summary of related work and a description of research methods in Sections 2 and 3. Sections 4 and 5 provide the results of the experiments and discussion on the results with conclusions, respectively.

2. **Related Work.** The evaluator discovers different usability problems depending on the usability evaluation method used [3], and another factor that can influence the usability problems found is the number and characteristics of users or evaluators participating in usability experiments [5-7] and expert reviews [8,9]. Hertzum and Jacobsen [3] coined the term evaluator effect to indicate the difference in problem detection and severity judgment by evaluators using the same usability evaluation method. Several previous studies have suggested the existence of an evaluator effect in usability evaluation. For example, a comparative usability evaluation study found low redundancy in the set of usability problems found by a team of experts evaluating the same interface [4]. Evaluators also differ in their judgment of the severity of the usability problem found [10,19]. According to the results of several previous studies, which reported that the usability problems discovered by the evaluators were very diverse, it can be expected that the characteristics of the evaluators can affect the description types of the usability problems when reporting. In this study, gender (male/female), language (Korean/English), and AR UI experience (high/low/none) were considered as variables of the evaluator's characteristics, and the effects of these on the evaluator's description type of usability problems were investigated.

Heuristic evaluation is one of the most commonly used usability evaluation methods. This method employs a group of experts who systematically evaluate a system's interfaces, utilizing well-known UI design principles called heuristics. Nielsen [11] proposed a set of ten general heuristics that are widely used for heuristic evaluation. Heuristic evaluation based on these ten general heuristics is easy to apply, but often fails to detect certain important usability problems. This is in part because Nielsen's heuristics are too general to evaluate systems with very specific properties. Because of this aspect, domain-specific heuristics have been developed to evaluate the UIs of newly developed systems with advanced technologies and complex interfaces rather than general heuristics. These two types of heuristics are commonly referred to in the literature as general heuristics and specific heuristics [11,12,20]. For example, especially in AR environments, the type of heuristic needs to be considered to apply heuristic evaluation. General heuristics, such as Nielsen's ten heuristics, do not include interface features such as finding, selecting and manipulating objects in 3D space. Similarly, input and output modalities can be fundamentally different from those addressed by conventional heuristics for AR interfaces, requiring a different evaluation approach [17]. Therefore, in this study, heuristic type was considered as a variable influencing the detection and description of usability problems based on heuristic evaluation, and two types of heuristic were used in the experiments: general heuristics and specific heuristics.

Augmented reality (AR) can visually augment information in the real physical world by incorporating additional information [13]. In other words, AR is a system that supplements the real world with virtual objects synthesized by a computer, and the real world and virtual objects coexist in the same space. Therefore, AR can enhance the interaction between the user and the interface through various modes of activity [14,15,21], and AR UIs may be classified according to the degree of interaction with the user. In this study, heuristic evaluation was conducted with two kinds of AR UIs that required two levels of interaction: high-level and low-level user interaction.

3. **Method.** In order to find which factors affect the evaluator's description types of usability problems when heuristic evaluation is applied to AR UIs, the experiments were conducted with a total of 46 participants.

3.1. **Experimental design.** Mixed effects design was applied to the experiments with five independent variables: heuristic type (general heuristics vs. specific heuristics), user interaction level (high vs. low), gender (male vs. female), language (Korean vs. English), and AR experience (high vs. low vs. none). Among 46 participants, 23 participants used general heuristics and the other 23 participants used specific heuristics (i.e., between-subject design), and every participant conducted tasks with two kinds of AR UIs that required high and low user interaction levels, respectively (i.e., within-subject design). Dependent variable is the evaluator's description type of usability problem (abstract type vs. concrete type). After heuristic evaluation was applied to AR UIs, the participants found usability problems and described them in the report form. Usability problems described by participants as evaluators were firstly classified into abstract and concrete usability problems by the judgment of two human-computer interaction experts, and then participant's description of usability problems was classified into two types: abstract type when the proportion of abstract usability problems out of all usability problems found is greater than 0.1, and concrete type in other cases.

3.2. **Augmented reality user interfaces.** For the experiments, AR-based user interfaces were prepared using Unity and Vuforia engines and implemented on a tablet PC (Galaxy tab), and participants were asked to evaluate these AR interfaces and discover usability problems while performing maintenance tasks using the AR interfaces. When performing the task, the user interaction level with the AR interfaces was designed to be high and low, and the language used was set to Korean and English, so all four types of AR user interfaces were prepared as shown in Figure 1.

In the case of AR UI to which high user interaction level is applied (see (a) and (c) in Figure 1), the participants have to perform tasks while actively finding a lot of information for each maintenance step, such as pressing a screen button or marker-based virtual button, whereas in the case of AR UI to which low user interaction level is applied (see (b) and (d) in Figure 1), the participants perform tasks passively only using information presented on the UI. The information provided on the AR UI is related to instructing the task of replacing a computer component such as a cooling fan and a memory card, and is displayed in Korean (see (c) and (d) in Figure 1) and English (see (a) and (b) in Figure 1), respectively.

3.3. **Participants.** A total of 46 participants participated in the experiments. They are 23 males and 23 females, and mostly college students in their 20s (average: 21.7 years old, standard deviation: 2.3 years old). In terms of language, 23 participants of Korean nationality used Korean to conduct the experiment, and the other 23 participants were multinational participants who communicated well in English and conducted the

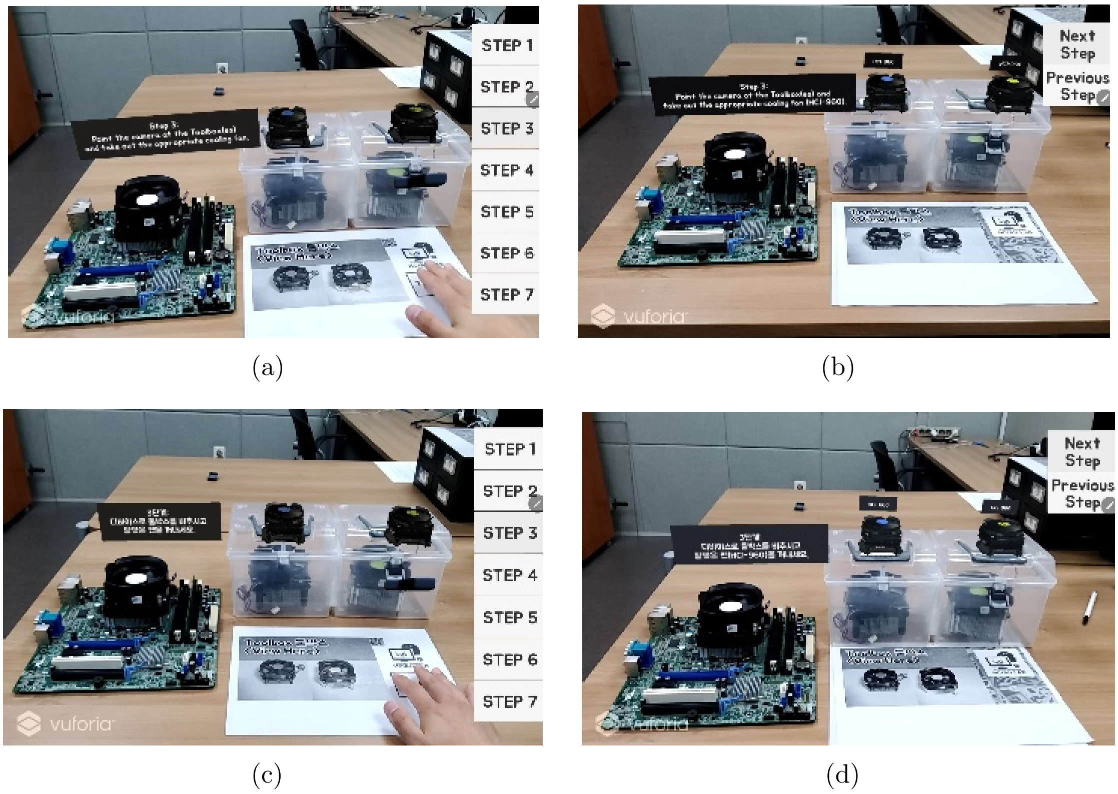

Figure 1. Augmented reality user interfaces for experiments: (a) High user interaction level (English version), (b) low user interaction level (English version), (c) high user interaction level (Korean version), and (d) low user interaction level (Korean version)

experiment using English. The AR experience was divided into three levels: high level when participants experienced more than once a week, low level when participants had experience but not more than once a week, and no experience at all. In this study, 14 participants belong to the high AR experience level, 18 participants belong to the low AR experience level, and the remaining 14 participants have no AR experience at all.

3.4. **Procedures.** 23 participants were randomly selected to use general heuristics and the remaining 23 participants used specific heuristics to find usability problems in AR UIs. At this time, it was considered that gender and language were distributed as evenly as possible. The general heuristics consisted of 23 heuristics at the level of general UI design principles, and the specific heuristics consisted of 110 heuristics at the level of the specific UI design guide. Participants perform a task to experience interaction with the AR UI. The task is to replace computer components, such as a cooling fan and a memory card, according to information provided by the AR UI. AR UIs were prepared in two forms according to two user interaction levels (high vs. low), and each participant conducted a total of two experiments using these two AR UIs respectively. The order of two experiments within a participant was counter-balanced among participants to avoid the possible order effects, and the wash-out period of at least 24 hours was given between two experiments. Therefore, there were a total of 92 experimental sessions (46 participants × 2 user interaction levels of AR UIs). Each of participants conducted the experiment in the following order: (1) The experiment, including purpose and procedure of the experiment, and heuristics and AR UIs, is introduced to a participant; (2) Training time is allowed to a participant to familiarize him/herself with the AR UI and the task; (3) The participant interacts with the AR UI to perform the task; (4) The participant finds usability problems of AR UI using the assigned heuristics and describes them freely in the prepared sheet.

4. **Results.** After the evaluator's description type of usability problem (abstract type vs. concrete type) was determined based on the proportion of abstract usability problems out of all usability problems found, the data for the dependent variable was coded as 1 for abstract type and 0 for concrete type to apply mixed effects logistic regression. Mixed effects logistic regression is used to model binary dependent variables in which the log odds of an outcome are modeled as a linear combination of independent variables in the presence of both fixed and random effects. In this study, the evaluator's description type of usability problem as the dependent variable shows binary outcomes (i.e., 1 for abstract type and 0 for concrete type), and the independent variables include 'participant' as a random factor and 'heuristic type (general vs. specific)', 'user interaction level (high vs. low)', 'gender (male vs. female)', 'language (Korean vs. English)' and 'AR experience (high vs. low vs. none)' as fixed factors.

As a result of mixed effects logistic regression, Table 1 shows the estimate of regression coefficient for the fixed factors, and 'user interaction level ($p = 0.0005$)' and 'language ($p = 0.0000$)' are significant factors to explain the log odds of the outcomes (i.e., the evaluator's description type of usability problem). The estimates of regression coefficients for 'user interaction level (low)' and 'language (Korean)' are '−8.57974' and '26.90297', respectively.

Table 1. Mixed effects logistic regression coefficients

Variables	Estimate	Std. error	z value	p-value
(Intercept)	−8.29684	3.16341	−2.623	0.00872***
Heuristic type (specific)	0.04078	2.71865	0.015	0.98803
User interaction level (low)	−8.57974	2.12678	−4.034	5.48E-05***
Gender (male)	−0.12348	2.69598	−0.046	0.96347
Language (Korean)	26.90297	5.32871	5.049	4.45E-07***
AR experience (low)	0.23051	3.70673	0.062	0.95041
AR experience (none)	−1.19513	3.143	−0.38	0.70376

Note: *** $p < 0.01$

The odds ratio values can be calculated from the estimates of regression coefficients through the exponential function, and are shown in Table 2. The odds ratio values for 'user interaction level (low)' and 'language (Korean)' are '0.000187874' and '4.8285E+11', respectively, and the other values are close to 1. The odds ratio value is usually interpreted according to whether it is greater than or less than 1, and if the odds ratio value of a factor equals 1, it cannot be considered as having a significant effect. From the fact that the odds ratio value for 'user interaction level (low)' is much smaller than 1, it can be interpreted that when evaluating the usability of AR UI that requires a low level of user interaction, the odds, which would be an abstract description type of usability problems, are much lower than when evaluating the usability of AR UI that requires a high level of user interaction. In addition, from the fact that the odds ratio value for 'language (Korean)' is much larger than 1, it can be interpreted that the odds, which would be an abstract description type of usability problems when using Korean, are much higher than when using English.

According to the analysis of mixed effects logistic regression based on the empirical data from the heuristic evaluation, the results can be summarized as follows. First, the heuristic type (general heuristic vs. specific heuristic) does not affect the evaluator's description type of usability problem (abstract type vs. concrete type). Second, the user interaction level (high vs. low) of AR UIs affects the evaluator's description type of usability problem (abstract type vs. concrete type) significantly, and it can be concluded that there is a tendency for high user interaction level of AR UIs to make evaluators describe usability

Table 2. Odds ratio and its confidence interval (CI)

Variables	Odds ratio	2.50% CI	97.50% CI
(Intercept)	0.000249304	5.06E-07	0.122879848
Heuristic type (specific)	1.041625354	0.005053012	214.7201256
User interaction level (low)	0.000187874	2.91E-06	0.01213997
Gender (male)	0.883843139	0.004482376	174.2778211
Language (Korean)	4.8285E+11	14057884.05	1.65846E+16
AR experience (low)	1.259245486	0.000880805	1800.284402
AR experience (none)	0.302663516	0.000639126	143.3288406

problems in an abstract way. Third, among the factors related to the evaluator's characteristics, such as gender, language and AR experience, only language (Korean vs. English) affects the evaluator's description type of usability problem (abstract type vs. concrete type) significantly, and gender and AR experience are not significant. It can be concluded that there is a tendency for the evaluators using Korean to describe usability problems in an abstract way. Especially, to the conclusion that high user interaction level of AR UIs makes evaluators describe usability problems abstractly, we can interpret that when a lot of cognitive resources such as attention are used for interaction with the UI, there is a tendency to use the abstract description method, which uses less cognitive resources when describing usability problems. In addition, from the conclusion that there is a tendency to describe usability problems in an abstract way when using Korean compared to English, it can be presumed that Korean has linguistic characteristics that have more connotative and abstract expressions than English.

5. **Conclusion.** From the experimental results, it can be concluded that user interaction level of UIs and charateristics of language used may lead to preference for abstract descriptions when describing usability problems. Therefore, we need to take account of these two factors to obtain the concrete description of usability problems when we use the heuristic evaluation to evaluate AR UIs. The contributions of this study are to empirically explore the factors affecting the evaluator's description type of usability problems when using the heuristic evaluation method, and to be able to give practitioners the implication that user interaction level and language should be considered as significant factor affecting the evaluator's description type of usability problems when planning the heuristic evaluation. However, in future research, more diverse factors that can affect the evaluator's description type of usability problems should be included in the study, and the study needs to be expanded to other usability evaluation methods other than the heuristic evaluation.

Acknowledgment. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1054050).

REFERENCES

[1] H. R. Hartson, T. S. Andre and R. C. Williges, Criteria for evaluating usability evaluation methods, *International Journal of Human-Computer Interaction*, vol.13, no.4, pp.373-410, 2001.

[2] T. S. Andre, H. R. Hartson, S. M. Belz and F. A. McCreary, The user action framework: A reliable foundation for usability engineering support tools, *International Journal of Human-Computer Studies*, vol.13, pp.107-136, 2001.

[3] M. Hertzum and N. E. Jacobsen, The evaluator effect: A chilling fact about usability evaluation methods, *International Journal of Human-Computer Interaction*, no.4, pp.183-204, 2003.

[4] R. Molich, M. R. Ede, K. Kaasgaard and B. Karyukin, Comparative usability evaluation, *Behaviour & Information Technology*, vol.23, no.1, pp.65-74, 2004.

[5] E. L.-C. Law and E. T. Hvannberg, Analysis of combinatorial user effect in international usability tests, *Conference on Human Factors in Computing Systems*, no.8, pp.9-16, 2004.

[6] J. Nielsen, Estimating the number of subjects needed for a thinking aloud test, *International Journal of Human-Computer Studies*, vol.41, pp.385-397, 1994.

[7] R. A. Virzi, Streamlining the design process: Running fewer subjects, *Proc. of the Human Factors Society Annual Meeting*, vol.34, no.4, pp.291-294, 1990.

[8] J. Dumas and J. Sorce, Expert reviews: How many experts is enough?, *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, vol.39, no.4, pp.228-232, 1995.

[9] A. Dutt, H. Johnson and P. Johnson, Evaluating evaluation methods, *Proc. of the Conference on People and Computers IX*, pp.109-121, 1994.

[10] M. Hassenzahl, Prioritizing usability problems: Data-driven and judgement-driven severity estimates, *Behaviour & Information Technology*, vol.19, no.1, pp.29-42, 2000.

[11] J. Nielsen, Heuristic evaluation, in *Usability Inspection Methods*, J. Nielsen and R. L. Mack (eds.), New York, NY, John Wiley & Sons, 1994.

[12] R. Parente Da Costa, E. D. Canedo, R. T. De Sousa, R. De Oliveira Albuquerque and L. J. García Villalba, Set of usability heuristics for quality assessment of mobile applications on smartphones, *IEEE Access*, vol.7, pp.116145-116161, 2019.

[13] M. Singh and M. P. Singh, Augmented reality interfaces, *IEEE Internet Comput.*, vol.17, no.6, pp.66-70, 2013.

[14] C. Kirner and T. G. Kirner, Virtual reality and augmented reality applied to simulation visualization, in *Simulation and Modeling: Current Technologies and Applications*, A. El Sheikh, A. T. Al Ajeeli and E. Abu-Taieh (eds.), IGI Global, 2008.

[15] R. T. Azuma, A survey of augmented reality, *Presence: Teleoperators & Virtual Environments*, vol.6, no.4, pp.355-385, 1997.

[16] J. Nielsen and R. Molich, Heuristic evaluation of user interfaces, *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp.249-256, 1990.

[17] A. Dünser and M. Billinghurst, Evaluating augmented reality systems, in *Handbook of Augmented Reality*, B. Furht (ed.), New York, NY, Springer, 2011.

[18] G. A. Wildenbos, M. W. M. Jaspers, M. P. Schijven and L. W. Dusseljee-Peute, Mobile health for older adult patients: Using an aging barriers framework to classify usability problems, *International Journal of Medical Informatics*, vol.124, pp.68-77, 2019.

[19] R. Khajouei, A. Ameri and Y. Jahani, Evaluating the agreement of users with usability problems identified by heuristic evaluation, *International Journal of Medical Informatics*, vol.117, pp.13-18, 2018.

[20] M. K. Othman, M. N. S. Sulaiman and S. Aman, Heuristic evaluation: Comparing generic and specific usability heuristics for identification of usability problems in a living museum mobile guide app, *Advances in Human-Computer Interaction*, vol.2018, pp.1-13, 2018.

[21] K. Ogata, S. Nakatani and S. Yamaguchi, A network camera system enabling long-distance use of augmented reality functionality using ARToolKit, *International Journal of Innovative Computing, Information and Control*, vol.17, no.2, pp.655-669, 2021.