

A User Study Trends in Augmented Reality and Virtual Reality Research

A qualitative study with the past three years of the ISMAR and IEEE VR conference papers

Si Jung Jun Kim

Human-Computer Interaction Lab
Digital Media, School of Visual Arts and Design
University of Central Florida, Orlando, FL 32826
sjkim@ucf.edu

Abstract—Augmented reality (AR) and virtual reality (VR) are becoming a part of everyday life with the advance of other technologies such as computer vision systems, sensing technologies, graphics, mobile computing, etc. Their primary goal is to help users achieve their goals effectively and efficiently with satisfaction. This paper describes the trends of how user studies have been incorporated into AR and VR papers published in two major conferences over the past three years. In addition, this paper presents implications on what needs to be taken into account when planning a user study in the field of AR and VR research.

Keywords—usability, evaluation, user study

I. INTRODUCTION

The virtual reality (VR) refers to systems that simulate the real world and augmented reality (AR) refers to systems that add virtual information to the real world. They are becoming a part of everyday life as computing devices become smaller, faster, and more ubiquitous. As this field gains popularity, there are many professional societies dealing with AR and VR research such as ISMAR (International Symposium on Mixed Augmented Reality), IEEE VR (Virtual Reality; formally VRAIS), ISUVR (International Symposium on Ubiquitous Virtual Reality), JVR (Joint Virtual Reality Conference), World Conference on Innovative Virtual Reality, etc. Their goal is to bring together AR/VR related research projects to a venue where students, faculty and practitioners meet together to share and discuss their projects aimed at helping users achieve their goals effectively and efficiently. This paper presents a qualitative study that investigated how AR and VR related research projects published in IEEE VR and ISMAR conferences over the past three years incorporated user based studies to meet their users' needs and increase the level of usability. The paper also provides implications that address how the concept of usability can be incorporated into the area of AR and VR research.

II. BACKGROUND

The field of AR and VR research is primarily for human use as it is augmenting or simulating the real world. User studies are crucial in AR and VR research as it represents how users respond to stimuli in the research and helps to

match more closely to user needs and requirements that improve the effectiveness or efficiency of the research [1]. Research has been done to address these. A set of evaluation techniques especially in the AR research was introduced [2]. It was a literature survey based on over hundred AR related papers that addresses user study approaches and methods as well as user evaluation types. A structured collection of usability design and evaluation guidelines were presented that synthesized information from a literature survey with many different sources[3]. Specific usability techniques for the VR research has been introduced [4, 5]. In the research, the practical significance of identifying and accommodating individual differences has been established across a number of fields of research. For instance, the significance of considering individual differences was emphasized to be taken into account in conducting research. These include user's experience levels, physical capabilities and limitations and technical aptitudes [6-9]. A tutorial about conducting experiments with human subjects in the field of AR and VR research were given in the 2004 IEEE VR conference and it is being held almost every year of the conference [10]. This implies the practical significance of integrating the user study into the field of AR and VR research.

III. RESEARCH DESIGN

A research design has been formed to address the question of how AR and VR research have been conducted to meet their users' needs. The main idea for the question was to review AR and VR research papers published in major AR and VR conferences and find a trend.

A two-factor between-subjects design was formed, where one factor is types of research papers and the other is different evaluation criteria. A total of 364 papers published in the ISMAR and IEEE VR conferences over the past three years from 2009 to 2011 were considered as the first independent variable shown in Table 1.

The ISMAR and IEEE VR have been chosen because these are the major societies contributing to in the field of AR and VR research. This year's IEEE VR proceeding was not considered because it is not yet published and ISMAR 2012 is not held. An interesting thing that we noticed in the preparation of the IEEE VR and ISMAR papers was that both societies have less publication numbers in 2010 compared to that of 2009 and 2011. A total of four review

criteria listed below have been established as the second independent variable.

1. Has the research been done an experiment with any type of participants?
2. Has the experiment been conducted with human subjects? If so, how many subjects were participated in the study?
3. Has the data been collected from the experiment analyzed using statistical analysis?
4. Does the study plan a future usability study?

TABLE I. NUMBER OF PAPERS PUBLISHED IN THE ISMAR AND IEEE VR CONFERENCES AND THE DISTRUBITON OF INDEPENDENT CODERS

Society	Year	Numbers of papers published	Distribution of Independent Coder
IEEE VR	2009	73	C1
IEEE VR	2010	69	C2
IEEE VR	2011	73	C3
Total		215	
ISMAR	2009	52	C4
ISMAR	2010	45	C5
ISMAR	2011	52	C6
Total		149	
Grand Total		364	6

Six individuals, who are familiar with AR and VR were recruited as independent coder and they reviewed AR and VR papers according to the review criteria. The independent coders were assigned to the six conference proceedings shown in Table I, where a letter C denotes a coder and numbers represent their id. There was neither repetition on reviewing papers in each proceeding nor alternating the independent coders for the proceedings.

IV. RESULTS AND DISCUSSION

A total of six independent coders whose age ranged from 22 years old and 28 years old ($M=24$, $SD= 2.7$) were recruited. All the independent coders have sufficient knowledge both in the area of AR and VR that they answered above four out of seven degree Likert scale, where the seven is extremely knowledgeable.

A. Experiments Conducted with Any Participants

The first analysis was to know how many experiments with any types of participants including the authors were conducted in AR and VR research over the past three years. The results were interesting in that the rate did not exceed 50% at least one year over the past three years in both AR and VR research. As shown in Figure 1, the VR papers conducted an experiment with any participants are 28.8% (21 out of 73) in 2009, 46.4% (32 out of 69) in 2010 and 46.68% (34 out of 73) in 2011. The ISMAR papers have a lower rate than that of the VR papers. Experiments conducted with any types of participants in the AR papers are 25.0% (13 out of 52) in 2009, 28.9% (13 out of 45) in 2010 and 36.5% (19 out of 52) in 2011. In terms of quantity, it seems like experiments were conducted more in the VR research rather than in the AR research, however the rates of the experiments conducted with participants in the AR research were increased more as the years go than that of the VR

research. Both AR and VR research have the same tendency that experiments in the fields are getting more increased whether or not they are for measuring user experience or system performance. It showed that conducting experiments in 2011 in the VR research has increased 17.8% compared to the rate in 2009. There was a substantial increase in the VR research from 2009 to 2010, an increase that total 17.6% during that time. There is the same trend in the AR research that about 11.5% has increased in 2011 compared to the rate in 2009.

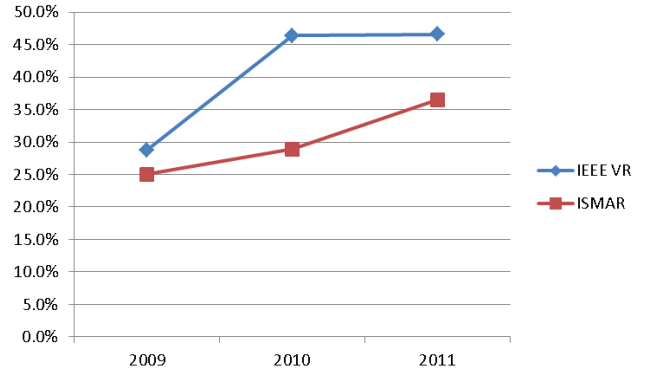


Figure 1. Experiments conducted in the papers of each year's proceeding.

B. User Studies Conducted with Users

This review was conducted to know whether AR and VR research projects have done with their users in their experiments. Figure 2 shows the results that the VR papers have conducted user studies with their users 27.0% (20 out of 73) in 2009, 39.1% (27 out of 69) in 2010 and 42.5% (31 out of 73) in 2011. The rates of the ISMAR papers are 23.1% (12 out of 52) in 2009, 28.9% (13 out of 45) in 2010 and 36.5% (19 out of 52) in 2011.

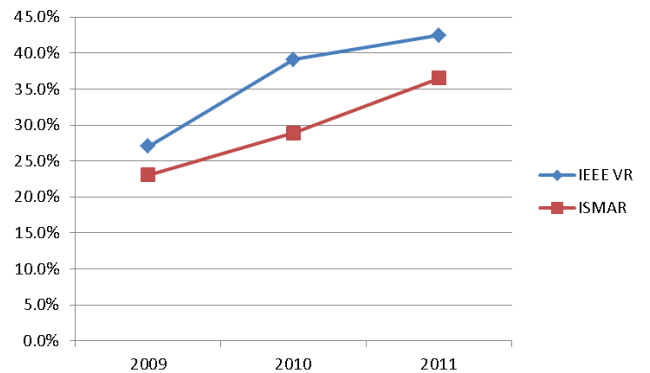


Figure 2. User studies conducted in AR and VR papers of each year's proceeding.

Similar to the results of the A. Experiments Conducted with Any Participants described in the previous section, the ISMAR papers have a lower rate than the rate of the VR papers in this evaluation category. It seems like the VR research conducts user studies with their users more than the

AR research. There is an interesting finding that shows that user studies have increased in both AR and VR research about 11.5% and 15.4% respectively over the past three years. The rate of the VR research in 2010 has increased dramatically (12.1%) compared to the rate of 2009 and the rate is continuously increasing in 2011. The rate of the AR research is also increasing gradually as years go, but it seems not very active showing the AR research tends to more focus on technological advances rather than testing their research with their users. However, the increased rates of the user studies in both AR and VR research imply the importance of incorporating the user study into the field of AR and VR research.

C. User Studies Conducted with over 24 Users

This review category was chosen to know how many research projects in AR and VR research were conducted with over 24 participants in generalizing their findings. Generally speaking, a sample size of 30 or more is considered to be large enough for the central limit theorem to take effect. The participant's number, 24 used in this review must be based on the types of population but the number 24 was used arbitrarily and assumed that the population is general and follows the normal distribution. Like the previous two analyses, this review also showed the same tendency that both AR and VR research tried to include more users to their user studies as Figure 3 shows.

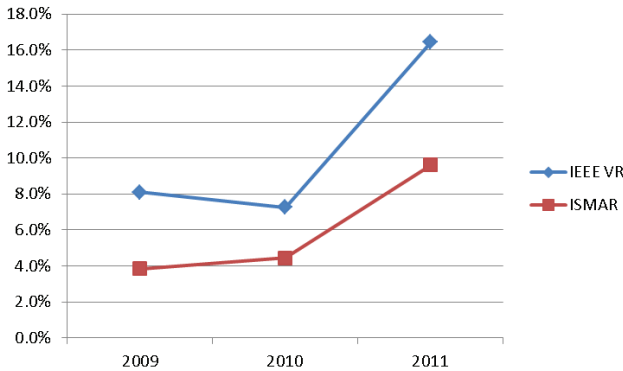


Figure 3. User studies conducted with over 24 users.

User studies conducted with over 24 participants in the VR research are 8.1% (6 out of 73) in 2009, 7.2% (5 out of 69) in 2010 and 16.4% (12 out of 73) in 2011. There was a big change in 2011 shown in Figure 3. The rates of the AR papers are 3.8% (2 out of 52) in 2009, 4.4% (2 out of 45) in 2010 and 9.6% (5 out of 52) in 2011. There was also substantial change in 2011 in the AR papers. It seems like both AR and VR research tend to include more users in their user studies (VR 8.3%, AR 5.8%), however the rate of conducting the user study with over 24 users was still low compared to the total papers accepted to each year's conference in each research area.

D. User Studies Planned in Future Works

The next review category investigated was how user studies planned their future works. It was to know how

much the authors considered incorporating the user study into the research with their users. The VR papers planned 29.7% (22 out of 73) in 2009, 55.1% (38 out of 69) in 2010 and 24.7% (18 out of 73) in 2011. The rates of the AR papers are 5.8% (3 out of 52) in 2009, 28.9% (13 out of 45) in 2010 and 15.4% (8 out of 52) in 2011. Figure 4 is a graphical representation of these rates.

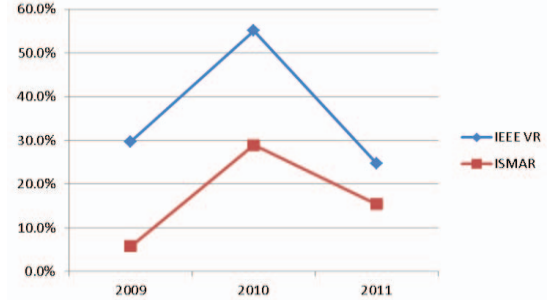


Figure 4. Rates of user studies planned as future works.

There was an interesting turning point shown in Figure 4 that the rate planned the user study for 2012 in 2011 both in AR and VR research is lower than the previous year's as 24.7% (18 out of 73) in the VR research and 14.4% (8 out of 52). It is lower than that of 2009, 29.7% (22 out of 73) and year 2010, 55.1% (38 out of 69).

The rates of the user studies planned for the next year and the actual rates of the user studies conducted in the year are plotted shown in Figure 5. As the graph shows, it appears that there is a positive correlation between the numbers of planned user studies for the next year and the numbers of user studies conducted in the year. As the rate of user study for future works increased, there were more papers that conducted the user study.

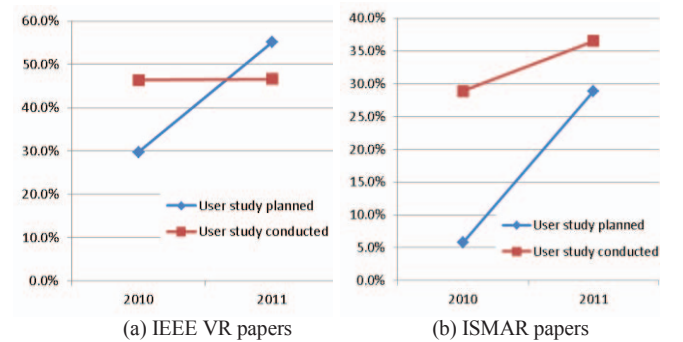


Figure 5. Relationships between user studies planned in the previous year and user studies conducted in the year.

The VR papers in 2009, about 29.7% (22 out of 73) planned the user study and there were 39.1% (27 out of 69) of papers conducted the user study in the 2010. In the 2010 VR research, about 55.1% (38 out of 69) planned the user study and there were 42.5% (31 out of 73) of papers performed the user study in 2011. In the AR research, about 5.8% (3 out of 52) planned the user study for 2010 in 2009 and there were 28.9% (13 out of 45) of papers performed the user study in 2010. In the 2010 AR research, about 28.9%

(13 out of 45) planned the user study and 36.5% (19 out of 52) of papers conducted the user study in 2011.

There is a tendency in the ISMAR papers that the AR research tends to plan user studies more and more as each year goes by. The user study rate planned in year 2009 was only 5.8% (3 out of 52) but the next year's was 28.9% (13 out of 45). The rate has been increased 23.1%, which is substantial. This implies that the AR research tends to incorporate the user study more in testing their effectiveness and efficiency with their users. A finding was that the rate of user studies conducted in the year was higher than the rate planned in the year before.

E. General Discussion

The four analyses investigated in this research discovered that user studies are getting more incorporated into both AR and VR research. In addition, the user studies conducted in AR and VR research were evaluated statistically regardless it was descriptive or inferential. For example, about 45.2% (33 out of 73) papers in the year of 2011 VR research were analyzed statistically, while about 32.9% (24 out of 73) papers were evaluated in 2009. It has been increased about 12.3% in three years. There was a similar trend in the AR research. About 9.6% (5 out of 52) of the AR papers in 2009 was evaluated statistically, but about 11.5% (6 out of 52%) of the AR papers was statistically evaluated in 2011. The rate has been increased to 1.9% but it is in the stage of increment. This implies that recent user studies are being analyzed based on statistical analysis. It was not investigated in this research whether the user data collected from the user study were based on empirical user studies, where actual users were involved with experiments and produced the data.

Considering the results of how many experiments and user studies were conducted both in AR and VR research, it seems like the VR research has more taken account into the user study than the AR research. In other words, the VR research appears to work with their users to test their effectiveness or efficiency, which is the primary goal of the user study and usability. A reason for this is technologies in the VR seem to have matured over the past two decades and researchers in the area tend to focus on investigating the effectiveness and efficiency of the technologies, while researchers in the AR tend to more focus on technological advancement such as in computer vision systems, sensing technologies, graphics, mobile computing, etc. rather than focusing on the evaluation of their research with their users.

Although types of the user studies conducted in AR and VR research over the past three years were not coded and identified in this investigation, it appears to be majority of the user studies were formative user studies that were conducted not of the design but for the design. The goal of the user studies were not coded and identified either in this research, but it seems like the majority of the user studies conducted in AR and VR papers were not aimed at testing the efficiency of the proposed research (e.g., doing the things in the right manner), but aimed at testing the effectiveness of proposed research (e.g., doing the right things).

It is suggested to start planning the user study based on these three areas: the user, goals and the context of use. The user is to identify that who is using the proposed the AR/VR research outcome. The goal is to find the purpose of the proposed research whether the research support what the user wants to do with it. The last area, the context of use is to find a usage situation to expect where and how the proposed AR and VR research outcomes will be used. In addition, the three usability evaluation criteria, efficiency, effectiveness and satisfaction are recommended to be taken into account in planning the user study.

- Efficiency: how much effort did users require to do their job with the proposed AR/VR solution?
- Effectiveness: how well did the proposed solution performed?
- Satisfaction: how well users are satisfied with the proposed solution?

When it comes to taking an experiment, it is recommended to conduct both the formative evaluation and summative evaluation. With regard to the formative evaluation, it is recommended to bring actual users into the design sessions and empower them to control the design sessions. Regarding the summative evaluation, it is suggested to include as many as participants, at least 24 participants, to the user study to increase the confidence in results and to generalize them. However considering the type of users must be an essential prerequisite given the context of use both in AR and VR research then conduct a pilot study before conducting a formal study.

V. CONCLUSION

This paper presented the trends on how user studies have been conducted in AR and VR research projects published in two major AR and VR conferences over the past three years. This paper addressed the characteristics of AR and VR research in terms of how user-based usability studies were incorporated into the fields. Both AR and VR research have a tendency of incorporating the user study into their research to test the effectiveness or efficiency of their research. It was discovered that the rate of conducting user studies in the past three years of AR and VR research was less than 50%. The rate tells us that AR and VR research need to work with their users more in their research.

Both AR and VR research are for human users and failure to understand users' needs and requirements of AR and VR research will result in user frustration or cause an unsatisfactory experience. Therefore, user-based evaluations, either formative or summative or both of them, need to be taken into account in the design and development of AR and VR research in order to produce more usable outcomes for their users. It is authors' belief that this paper will inspire people to incorporate the user study into their research and thus increasing the usability of their outcomes.

This study is limited in considering only the three years of AR and VR papers published in the ISMAR and IEEE VR. Other limitations are involving with the low number of independent coders and all the papers were not reviewed by the independent coders. The natural extension of this study is to consider more AR and VR papers published in other AR

and VR societies and to review them with a between-subjects design with more independent coders.

ACKNOWLEDGMENT

1. Gabbard, J.L. and J.E. Swan, *Usability Engineering for Augmented Reality: Employing User-based Studies to Inform Design*. IEEE Transactions on Visualization and Computer Graphics, 2008. **14**.
2. Dünser, A., R. Grasset, and M. Billinghurst, *A survey of evaluation techniques used in augmented reality studies*, in *ACM SIGGRAPH ASIA2008*: Singapore.
3. Gabbard, J. and D. Hix, *Researching Usability Design and Evaluation Guidelines for Augmented Reality (AR) Systems*, 2001.
4. Hix, D. and J. Gabbard, *Usability Engineering of Virtual Environments*. 2002.
5. Bowman, D., J. Gabbard, and D. Hix, *A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods*. Presence: Teleoperators and Virtual Environments, 2002. **11**(4): p. 404-424.

We are grateful to the independent coders for sharing their time and coding the papers. Special thanks to H. Kim, who provided useful resources and constructive feedbacks. We thank P Silva, J. Valera and B. Jamshidi for reviewing the manuscript.

REFERENCES

6. Egan, D.E., *Individual differences in human-computer interaction*. Handbook of Human-Computer Interaction, ed. M. Helander. 1988, Amsterdam: Elsevier Science Publishers.
7. Chen, C., M. Czerwinski, and R. Macredie, *Individual Differences in Virtual Environments- Introduction and overview*. Journal of the American Society for Information Science, 2000. **51**(6): p. 499-507.
8. Borgman, C.L., *All users of information retrieval systems are not created equal: An exploration into individual differences*. Information Processing & Management, 1989. **25**: p. 237-251.
9. Summers, V.A., et al. *Calibration for augmented reality experimental testbeds*; . in *Symposium on Interactive 3D graphics* 1999.
10. Swan, J.E., et al. *Tutorial: Conducting Human-Subject Experiments with Virtual and Augmented Reality*. in *IEEE Virtual Reality*. 2004. Chicago, IL, USA.