Jessie Y. C. Chen · Gino Fragomeni (Eds.)

LNCS 10909

Virtual, Augmented and Mixed Reality

Interaction, Navigation, Visualization, Embodiment, and Simulation

10th International Conference, VAMR 2018
Held as Part of HCI International 2018
Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part I

Part I

HCI 2018 INTERNATIONAL

Springer

Lecture Notes in Computer Science 10909

More information about this series at http://www.springer.com/series/7409

Jessie Y. C. Chen · Gino Fragomeni (Eds.)

Virtual, Augmented and Mixed Reality

Interaction, Navigation, Visualization, Embodiment, and Simulation

10th International Conference, VAMR 2018
Held as Part of HCI International 2018
Las Vegas, NV, USA, July 15–20, 2018
Proceedings, Part I

Springer

Editors
Jessie Y. C. Chen
US Army Research Laboratory
Aberdeen Proving Ground, MD
USA

Gino Fragomeni
US Army Research Laboratory
Orlando, FL
USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-91580-7 ISBN 978-3-319-91581-4 (eBook)
https://doi.org/10.1007/978-3-319-91581-4

Library of Congress Control Number: 2018944280

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

The 20th International Conference on Human-Computer Interaction, HCI International 2018, was held in Las Vegas, NV, USA, during July 15–20, 2018. The event incorporated the 14 conferences/thematic areas listed on the following page.

A total of 4,373 individuals from academia, research institutes, industry, and governmental agencies from 76 countries submitted contributions, and 1,170 papers and 195 posters have been included in the proceedings. These contributions address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The contributions thoroughly cover the entire field of human-computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas. The volumes constituting the full set of the conference proceedings are listed in the following pages.

I would like to thank the program board chairs and the members of the program boards of all thematic areas and affiliated conferences for their contribution to the highest scientific quality and the overall success of the HCI International 2018 conference.

This conference would not have been possible without the continuous and unwavering support and advice of the founder, Conference General Chair Emeritus and Conference Scientific Advisor Prof. Gavriel Salvendy. For his outstanding efforts, I would like to express my appreciation to the communications chair and editor of *HCI International News*, Dr. Abbas Moallem.

July 2018 Constantine Stephanidis

HCI International 2018 Thematic Areas and Affiliated Conferences

Thematic areas:

- Human-Computer Interaction (HCI 2018)
- Human Interface and the Management of Information (HIMI 2018)

Affiliated conferences:

- 15th International Conference on Engineering Psychology and Cognitive Ergonomics (EPCE 2018)
- 12th International Conference on Universal Access in Human-Computer Interaction (UAHCI 2018)
- 10th International Conference on Virtual, Augmented, and Mixed Reality (VAMR 2018)
- 10th International Conference on Cross-Cultural Design (CCD 2018)
- 10th International Conference on Social Computing and Social Media (SCSM 2018)
- 12th International Conference on Augmented Cognition (AC 2018)
- 9th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management (DHM 2018)
- 7th International Conference on Design, User Experience, and Usability (DUXU 2018)
- 6th International Conference on Distributed, Ambient, and Pervasive Interactions (DAPI 2018)
- 5th International Conference on HCI in Business, Government, and Organizations (HCIBGO)
- 5th International Conference on Learning and Collaboration Technologies (LCT 2018)
- 4th International Conference on Human Aspects of IT for the Aged Population (ITAP 2018)

Conference Proceedings Volumes Full List

1. LNCS 10901, Human-Computer Interaction: Theories, Methods, and Human Issues (Part I), edited by Masaaki Kurosu
2. LNCS 10902, Human-Computer Interaction: Interaction in Context (Part II), edited by Masaaki Kurosu
3. LNCS 10903, Human-Computer Interaction: Interaction Technologies (Part III), edited by Masaaki Kurosu
4. LNCS 10904, Human Interface and the Management of Information: Interaction, Visualization, and Analytics (Part I), edited by Sakae Yamamoto and Hirohiko Mori
5. LNCS 10905, Human Interface and the Management of Information: Information in Applications and Services (Part II), edited by Sakae Yamamoto and Hirohiko Mori
6. LNAI 10906, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris
7. LNCS 10907, Universal Access in Human-Computer Interaction: Methods, Technologies, and Users (Part I), edited by Margherita Antona and Constantine Stephanidis
8. LNCS 10908, Universal Access in Human-Computer Interaction: Virtual, Augmented, and Intelligent Environments (Part II), edited by Margherita Antona and Constantine Stephanidis
9. LNCS 10909, Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation (Part I), edited by Jessie Y. C. Chen and Gino Fragomeni
10. LNCS 10910, Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry (Part II), edited by Jessie Y. C. Chen and Gino Fragomeni
11. LNCS 10911, Cross-Cultural Design: Methods, Tools, and Users (Part I), edited by Pei-Luen Patrick Rau
12. LNCS 10912, Cross-Cultural Design: Applications in Cultural Heritage, Creativity, and Social Development (Part II), edited by Pei-Luen Patrick Rau
13. LNCS 10913, Social Computing and Social Media: User Experience and Behavior (Part I), edited by Gabriele Meiselwitz
14. LNCS 10914, Social Computing and Social Media: Technologies and Analytics (Part II), edited by Gabriele Meiselwitz
15. LNAI 10915, Augmented Cognition: Intelligent Technologies (Part I), edited by Dylan D. Schmorrow and Cali M. Fidopiastis
16. LNAI 10916, Augmented Cognition: Users and Contexts (Part II), edited by Dylan D. Schmorrow and Cali M. Fidopiastis
17. LNCS 10917, Digital Human Modeling and Applications in Health, Safety, Ergonomics, and Risk Management, edited by Vincent G. Duffy
18. LNCS 10918, Design, User Experience, and Usability: Theory and Practice (Part I), edited by Aaron Marcus and Wentao Wang

19. LNCS 10919, Design, User Experience, and Usability: Designing Interactions (Part II), edited by Aaron Marcus and Wentao Wang
20. LNCS 10920, Design, User Experience, and Usability: Users, Contexts, and Case Studies (Part III), edited by Aaron Marcus and Wentao Wang
21. LNCS 10921, Distributed, Ambient, and Pervasive Interactions: Understanding Humans (Part I), edited by Norbert Streitz and Shin'ichi Konomi
22. LNCS 10922, Distributed, Ambient, and Pervasive Interactions: Technologies and Contexts (Part II), edited by Norbert Streitz and Shin'ichi Konomi
23. LNCS 10923, HCI in Business, Government, and Organizations, edited by Fiona Fui-Hoon Nah and Bo Sophia Xiao
24. LNCS 10924, Learning and Collaboration Technologies: Design, Development and Technological Innovation (Part I), edited by Panayiotis Zaphiris and Andri Ioannou
25. LNCS 10925, Learning and Collaboration Technologies: Learning and Teaching (Part II), edited by Panayiotis Zaphiris and Andri Ioannou
26. LNCS 10926, Human Aspects of IT for the Aged Population: Acceptance, Communication, and Participation (Part I), edited by Jia Zhou and Gavriel Salvendy
27. LNCS 10927, Human Aspects of IT for the Aged Population: Applications in Health, Assistance, and Entertainment (Part II), edited by Jia Zhou and Gavriel Salvendy
28. CCIS 850, HCI International 2018 Posters Extended Abstracts (Part I), edited by Constantine Stephanidis
29. CCIS 851, HCI International 2018 Posters Extended Abstracts (Part II), edited by Constantine Stephanidis
30. CCIS 852, HCI International 2018 Posters Extended Abstracts (Part III), edited by Constantine Stephanidis

http://2018.hci.international/proceedings

10th International Conference on Virtual, Augmented and Mixed Reality

The full list with the Program Board Chairs and the members of the Program Boards of all thematic areas and affiliated conferences is available online at:

http://www.hci.international/board-members-2018.php

HCI International 2019

The 21st International Conference on Human-Computer Interaction, HCI International 2019, will be held jointly with the affiliated conferences in Orlando, FL, USA, at Walt Disney World Swan and Dolphin Resort, July 26–31, 2019. It will cover a broad spectrum of themes related to Human-Computer Interaction, including theoretical issues, methods, tools, processes, and case studies in HCI design, as well as novel interaction techniques, interfaces, and applications. The proceedings will be published by Springer. More information will be available on the conference website: http://2019.hci.international/.

General Chair
Prof. Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
E-mail: general_chair@hcii2019.org

http://2019.hci.international/

Contents – Part I

Embodiment, Communication and Collaboration in VAMR

Education, Training and Simulation

Contents – Part II

VAMR in Psychotherapy, Exercising and Health

Virtual Reality for Cultural Heritage, Entertainment and Games

Industrial and Military Applications

Interaction, Navigation and Visualization in VAMR

Determining Which Touch Gestures Are Commonly Used When Visualizing Physics Problems in Augmented Reality

Marta del Rio Guerra[1,2], Jorge Martín-Gutiérrez[2], Raúl Vargas-Lizárraga[1], and Israel Garza-Bernal[1(✉)]

[1] Universidad de Monterrey, Garza García, NL, Mexico
{marta.delrio,raul.vargas,gustavo.garzab}@udem.edu
[2] Universidad de La Laguna, Tenerife, Spain
jmargu@ull.edu.es

Abstract. Touch gestures can be a very important aspect when developing mobile applications with enhanced reality. The main purpose of this research was to determine which touch gestures were most frequently used by engineering students when using a simulation of a projectile motion in a mobile AR application. A randomized experimental design was given to students, and the results showed the most commonly used gestures to visualize are: zoom in "pinch open", zoom out "pinch closed", move "drag" and spin "rotate".

Keywords: Augmented reality · Physics · Education · Usability
Touch gestures

1 Introduction

Augmented Reality (AR), as defined by Azuma, is an interactive system presented in real time that combines real and virtual realities registered in three dimensions [1]. Klopfer and Squire state that AR is "a real-world context dynamically overlaid with coherent location or context sensitive virtual information" [2]. Kipper and Rampolla mention in their book "Augmented Reality: An Emerging Technologies Guide to AR" that it is a variation of a Virtual Environment (VE). Our purpose is to use "digital or computer-generated information: whether they are images, audio, video, and touch or haptic sensations and overlay them in a real-time environment" [3]. Briefly, virtual information becomes real by using AR.

2 Related Work

Searching for a theoretical framework, Martín-Gutiérrez et al. introduced an AR application that worked with an AR book called AR-Dehaes, to enhance the acquisition of engineering students' spatial abilities. They made an experiment with undergraduate students from the Mechanical Engineering program and determined that the application

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 3–12, 2018.
https://doi.org/10.1007/978-3-319-91581-4_1

had a positive impact in relation to easiness, learnability, controllability and intrusiveness [5].

Regarding the scope of physics, Cai et al. created a system for a convex imaging experiment that transmitted an interactive AR video by means of the internet, allowing (eighth-grade) students to interact with three-dimensional models in the technologically advanced environment. The results showed that the main grades from the experimental group, were better than those belonging to the control group, but there was no significant difference in the main scores of the subsequent tests [4].

A mobile collaborative AR simulation system of elastic collisions developed by Lin et al. showed a notorious improvement in the learning outcomes of undergraduate students compared to those who studied in the traditional manner. They identified three different knowledge construction behaviors among students: construction of problem space, construction of conceptual space and construction of relations between conceptual and problem space. The results showed a significant difference in the subsequent test grades in both, the control and experimental groups and demonstrated that studying with the AR Physics system denoted better learning results than those achieved by students using the traditional method [6].

Akçayır et al. also investigated mobile AR applications in science laboratories by including undergraduate students in an experimental controlled group, which showed an upgrading in laboratory skills and attitudes when using improved physics laboratories instead of employing printed laboratory manuals. They developed five applications for five different physics experiments (water electrolysis, Ohm's law, Wheatstone bridge, Kirchoff's law and three phase transformer connections) and gave students attitude questionnaires, pre-tests, subsequent tests and interviews [7].

Martin-Gonzalez et al. developed an AR system with a Kinect sensor to teach Euclidean vectors in College institutions, in physics and mathematics courses, which enabled the understanding of concepts such as magnitude, direction, orientation and operations related to vectors (addition, subtraction and cross product). The user's hand positions were the final position for the vectors, and the origin was located on his chest [8].

Another work was done by Kaufmann and Meyer, who developed an enlarged reality application for mechanics education to simulate physical experiments. High School students were allowed to build their own physics experiments in a three-dimensional virtual world. As they mentioned in their research, the applicability of the app was for students aged 12 to 18, depending on the specific physics course; the researchers also thought that the app could be used for basic college courses [9].

An investigation on the use of a haptic augmented simulation of gravitational forces between the sun and the earth for undergraduate students, made by Civelek and Kemal explored the effects of this technology on the students' results and their attitudes towards physics. To determine those effects, they used an experimental and controlled group and the outcomes showed a positive effect on the students' grades [10].

Enyedy et al. invented a project named LLP (Learning Physics through Play) to teach physics topics to 6 to 8 years old students with activities using augmented reality [11]. The results showed that students developed a conceptual understanding of speed, force, net force, friction and motion at an earlier age using the AR technology.

An empirical study made by Ibanez et al. used an augmented reality simulator in a physics course. This simulator was used by 112 students to interact with magnetic fields in 3D. The results showed that AR is a technology that has a positive impact on the students' motivation [15].

Buchanan et al. developed a toolkit with AR and rigid body simulations to teach chain reactions and physical contraptions by means of a game. This toolkit provided an imitation of a chain reaction based on augmented objects that were visualized using different indicators. It was a success because the results showed that this toolkit helped participants learn Newton's forces since they all finished the simulation correctly [16].

A study made by Pasareti et al., which was conducted in a high school and dealt with augmented reality in education, showed that participants of an AR group obtained the following results: 60% had the best marks, 38% had average grades and only 2% had worse results compared to those obtained in a previous test; and the results of the control group showed 27% had better grades, 42% had average results and 31% had worse grades. After analyzing those results, they came to the conclusion that AR is a very good technology to be implemented in addition to textbooks [17].

Freitas and Campos in their study about SMART, an educational system with augmented reality, to teach 2nd grade students, presented the results of an experiment which was organized in three different elementary schools. There, 32 girls and 22 boys ranging from 7 to 8 years old, were divided in two groups: an experimental group that worked with SMART and a control group that studied with the traditional methodology. The results, based on the Portuguese educational system, were divided into three groups: poor students, average students and good students. The first group showed 34% better learning results which derived from the use of SMART, the second group showed 62% and the last one 33% [18].

As it was previously expressed, AR improves learning in physics topics, so we planned, worked and developed an AR mobile application to test which touch gestures are frequently used by college engineering students when using simulations of projectile motion problems in mobile applications. A collection of qualitative and quantitative tools was employed, such as ethnography observations and open interviews.

3 Methodology

3.1 Participants

The participants were all college students enrolled in Physics I course at Universidad de Monterrey in Mexico, during the Fall semester of 2017. With a population of 375 students, a confidence level of 95% and an interval of 15%, a sample of 26 students was obtained and two groups were formed according to their level of spatial intelligence: an outstanding and an average one. The following question was asked at the beginning of the experiment in order to establish the amount of spatial intelligence each student had: "When a physics problem is presented to you, how easily can you visualize it in your mind?" In this research work, 18 male and 8 female students were included, whose ages ranged from 17 to 20 years old, 12 of them showed an outstanding spatial intelligence

(9 males and 3 females) and 14 showed an average spatial intelligence (9 males and 5 females).

3.2 AR Physics Application

In order to carry out our research, we looked for an app with augmented reality that included tactile gestures and a simulation belonging to a physics subject. But nothing was found in the applications markets. So, a mobile application had to be built to carry out the tests on the participants. This system was created to help physics students improve their classroom participation and to enable student-technology interactions.

A mobile AR application was developed using Unity version 2017.2 with Vuforia version 6.5 and the Java Development Kit version 8, in order to export the app to Android System. The Vuforia package was imported to use the AR Camera instead of the main camera [12]. First, a marker was used to be recognized by the AR Camera in order to present the 3D model in augmented reality as shown in Fig. 1. Next, the LeanTouch package was downloaded to implement the mobile gestures for the 3D model. The following tactile gestures were programmed: moving, zooming in, zooming out and rotating, as shown in Figs. 1, 2, 3 and 4.

Fig. 1. Move simulation to any place of the screen with a touch gesture.

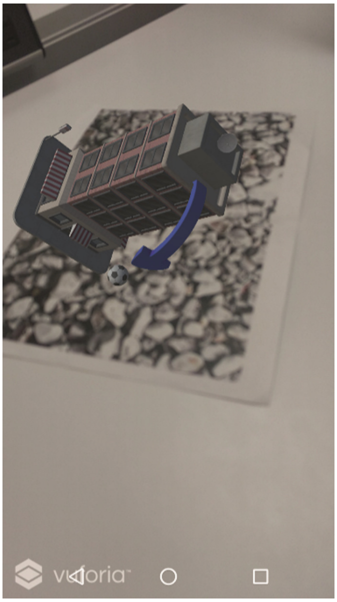


Fig. 2. Rotate simulation with a touch gesture.

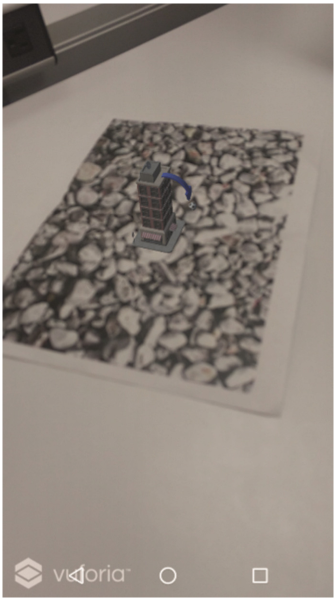


Fig. 3. Zoom-out simulation with a touch gesture.

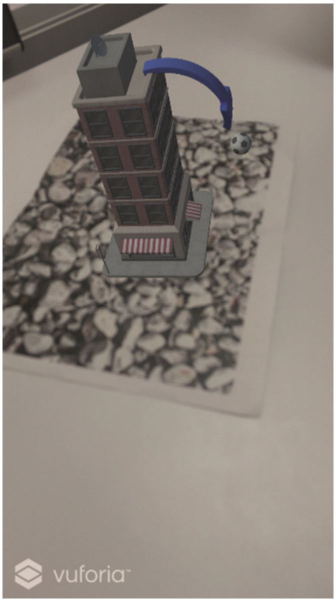


Fig. 4. Zoom-in simulation with a touch gesture.

As we can see in the previous images, the parabolic problem is working in augmented reality by applying the gesture to the 3D object (formed by the building and the blue arrow indicating where the ball will fall).

Finally, in order to export the application for Android System, an Android Studio 3.0 was downloaded with the SDK, in order to have the API's for each Android Version. The mobile app was able to work with Android 4.0 onwards, to the most up-to-date version (Android 7.1.1).

3.3 Procedure

The experiment was prepared individually for each participant. First, a tutorial was shown with the touch gestures that were implemented in the app in order to let students know which ones could be used. Then, we presented to them the following projectile motion problem: "From the top of a 15 meters-high tower, a projectile is launched with an initial velocity of 5 m/s and an elevation of 40°. Determine the horizontal distance travelled by the object". Then, we asked them to answer the following questions with the app.

- Question 1: What touch gesture would you employ to see the ball at a closer distance?
- Question 2: What touch gesture would you use to see the whole simulation?
- Question 3: What touch gesture would you use to move the simulation to the left?
- Question 4: What touch gesture would you practice to look at the horizontal distance the ball must travel?

Finally, we gave students an open interview in order to know what they had achieved when they finished the four tasks. The following questions were asked to find out what they thought about using AR in an application to be employed in the physics course.

1. Did the AR app help you understand the problem?
2. Did you understand the problem better by reading about it or by using the app?
3. Would the app motivate you to learn the material of the physics' course?

4 Outcome and Discussions

4.1 Touch Gestures Results

After carrying out the experiment, the touch gestures were obtained and analyzed. The results are presented in Table 1.

Table 1. Touch gestures results of the research

Question	Gesture	Times that the touch gesture was used	% of used gesture in each question
Question 1	Zoom-in with two fingers	24	92.30
	Zoom-in with double tap	2	7.70
Question 2	Zoom-out with two fingers	25	96.15
	Zoom-out with double tap	1	3.85
Question 3	Swipe	17	65.38
	Keep pressed and swipe	9	34.62
Question 4	Swipe to rotate	19	73.07
	Spin with two fingers	7	26.92

4.2 Students' Achievements

The data collected from the questions in order to obtain the students' results is presented in Table 2. Most students commented that the app helped them understand the problem we presented to them; They also commented that the best way to understand the physics' problems is by interacting with the simulator.

Table 2. Students' commentaries on their success.

Comment	Number of persons
It helps me because I can see it in 3D	20
I will not have grammar problems to understand the text	4
I will not have any problems to sketch a traditional 3D drawing	3
I will not have to waste time outlining the problem	1

4.3 Discussion

The majority of the students' commentaries on their results, it expressed was that the app fulfilled its purpose. The participants showed a positive reaction to the application when it was presented to them and they were interested in using it during the course, to get a better understanding of the physics problems that they were working with. To answer any problem and to be able to solve it, one must understand and outline it, so it can analyzed. However, there are cases in which students are unable to visualize it and delineate it. So, our application will help students solve both problems, outstanding spatial intelligence and average spatial intelligence.

With regards to the application, we must say that we are encouraged to continue working with it because of the participants' positive opinions since most of the students said that they would like to have augmented reality included as part of the curricula, to help them solve problems. That would save them time to sketch any problem presented to them and they would interact with it by means of gestures, in order to understand it.

In a future research work, we would like to add some exercises covering all the topics presented in the Physics 1 course, at Universidad de Monterrey, so that it can be integrated as a class tool.

4.4 Restrictions of the Study

When the experiment was carried out, there were two restrictions in our research work. The first one was the mobile application's development because it was impossible to program all the gestures shown in Table 1. The "LeanTouch" library of tactile gestures used for models working in augmented reality, only offers the gestures of zooming in, zooming out, rotate with two fingers, and swipe to move the simulation. To solve those problems, we gave the participants four questions, and also gave them instructions to refrain from touching the screen. Then, we asked them to pretend they made motions in the air. Finally, we let the participants interact with the application so that they could be in contact of the programmed gestures.

The second restriction was the size of the sample: it was not as big as we had planned at the beginning. One of the requirements was that students should be enrolled in the Physics 1 course and there was a problem to contact students who wanted to participate in our research. To find a solution and get an adequate sample to test our research, we spoke to the person in charge of the Physics Department of the University of Monterrey, who teaches the subject to three groups. He supported us by allowing us to take students to the classroom where we had the prepared equipment for the test.

5 Conclusions

After having carried out this research project and obtaining the results of the test, we counted the gestures used in each task and determined which one was used most frequently in each of the following classification: movement, zoom out, zoom in, and rotation.

The classification of the two gestures in order to approach the simulation to the application, the most frequently used gestures were “pinch open” (92.3%), “pinch closed” (96.15%), “drag” (65.38%), and rotate (73.07%).

In the classification of the two gestures for rotation of the simulator in the application, the most frequently used was the “Swipe” gesture showing a total of 19 times against 7 which was the gesture of “Spin with two fingers”. This was the classification with the most difficulty in the test for the participants to be able to successfully complete the task defined in the fourth question.

The results obtained in this study show that it is easy for physics students to perform a certain gesture to complete a task in the application. We verified that the app is easier to use by the type of user to whom it is being directed. According to Joan, in his study about enhancing education through augmented reality, students are able to enrich themselves by the use of augmented reality through electronic devices, since they are capable of relying on this tool anywhere and anytime; the pedagogical effects of using it in classroom were studied and it is stated that the students, who experience the use of augmented reality, developed better thinking skills [13]. It is deduced from this study that the development of the application would be a helpful tool for students who need a better understanding of physics problems and that tactile gestures are the right way to interact with augmented reality. Finally, with this app, an interactive education can be integrated in the near future into the physics course; Lee said in his study, that AR is a great way to make education interactive, since this technology allows “educational environments to be more productive, pleasurable, and interactive than ever before” [14].

References

1. Azuma, T.: A survey of augmented reality. Presence **6**(4), 355–385 (1997)
2. Klopfer, E., Squire, K.: Environmental detectives the development of an augmented reality platform for environmental simulations. Educ. Technol. Res. Dev. **56**(2), 203–228 (2008)
3. Kipper, G., Rampolla, J.: Augmented Reality: An Emerging Technologies Guide to AR. Elsevier, Waltham (2013)
4. Cai, S., Chaing, F.K., Wang, X.: Using the augmented reality 3D technique for a convex imaging experiment in a physics course. Int. J. Eng. Educ. **29**(4), 856–865 (2013)
5. Martín-Gutiérrez, J., Saorín, J., Contero, M., et al.: Design and validation of an augmented book for spatial abilities development in engineering students. Comput. Graph. **34**(1), 77–91 (2010)
6. Lin, T.-J., Duh, H.B.-L., Li, N., Wang, H.-Y., Tsai, C.-C.: An investigation of learners’ collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system. Comput. Educ. **68**, 314–321 (2013)
7. Akçayır, M., Akçayır, G., Pektaş, H.M., Ocak, M.A.: Augmented reality in science laboratories: the effects of augmented reality on university students’ laboratory skills and attitudes toward science laboratories. Comput. Hum. Behav. **57**, 334–342 (2016)
8. Martin-Gonzalez, A., Chi-Poot, A., Uc-Cetina, V.: Usability evaluation of an augmented reality system for teaching Euclidean vectors. Innov. Educ. Teach. Int. **53**(6), 627–636 (2015)
9. Kaufmann, H., Meyer, B.: Simulating educational physical experiments in augmented reality. In: ACM, SIGGRAPH Asia 2008. Association for Computing Machinery (2008)

10. Civelek, T., Kemal, M.: Effects of a haptic augmented simulation on K-12 students' achievement and their attitudes towards physics. Eurasia J. Math. Sci. Technol. Educ. **10**(6), 565–574 (2014)
11. Enyedy, N., Danish, J.A., Delacruz, G., Kumar, M.: Learning physics through play in an augmented reality environment. Comput. Support. Learn. **7**, 347 (2012)
12. Vuforia 6.5. https://developer.vuforia.com/downloads/sdk
13. Joan, R.: Enhancing education through mobile augmented reality. J. Educ. Technol. **11**(4), 8 (2015)
14. Lee, K.: Augmented reality in education and training. TechTrends **56**(2), 19 (2012)
15. Ibanez, M., De Castro, A., Kloos, C.: An empirical study of the use of an augmented reality simulator in a face-to-face physics course. In: IEEE 17th International Conference on Advanced Learning Technologies. Institute of Electrical and Electronics Engineers (2017)
16. Buchanan, P., Seichter, H., Billinghurst, M., Grasset, R.: Augmented reality and rigid body simulation for edutainment. In: 17th Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery (2014)
17. Pasareti, O., Hajdú, H., Matuszka, T., et al.: Augmented reality in education. In: INFODIDACT (2011)
18. Freitas, R., Campos, P.: SMART: a SysteM of augmented reality for teaching 2nd grade students. In: HCI 2008 Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction, vol 2, pp. 27–30 (2008)

Element Selection of Three-Dimensional Objects in Virtual Reality

Dylan Fox[1(✉)], Sophie So Yeon Park[1], Amol Borcar[2], Anna Brewer[3], and Joshua Yang[3]

[1] Center for Augmented Cognition and School of Information, University of California, Berkeley, USA
{dylan.r.fox,sophie.park}@berkeley.edu

[2] Haas School of Business, University of California, Berkeley, USA
amolborcar@berkeley.edu

[3] Virtual Reality at Berkeley, University of California, Berkeley, USA
https://goo.gl/D8HaLQ

Abstract. The manipulation of three dimensional objects is vital to fields such as engineering and architecture, but understanding 3D models from images on 2D screens takes years of experience. Virtual reality offers a powerful tool for the observation and manipulation of 3D objects by giving its users a sense of depth perception and the ability to reach through objects. To understand specific pain points in 2D CAD software, we conducted interviews and a survey of students and professionals with experience using CAD software. We narrowed in on the ability to select interior or obscured elements, and created a VR prototype allowing users to do so. Our usability tests found that compared to 2D software, VR was easier to use, more intuitive, and less frustrating, thought slightly more physically uncomfortable. Finally, we created a set of recommendations for VR CAD programs around action feedback, environmental context, and the necessity of a tutorial.

Keywords: Virtual reality · Usability · 3D modeling
Element selection · Computer-Aided Design (CAD)

1 Introduction

Virtual Reality (VR) has long been discussed for its potential to revolutionize business processes and activities ranging from video games to educational programs. Until recently, VR technology has been limited in its utility due to the cost and quality of video rendering, optics, and motion tracking technology. However, recent advancements in both software and hardware are vastly expanding the capabilities of VR programs. Several industries have taken notice and are attempting to integrate VR technology into their existing processes.

Engineering in particular stands to benefit from the introduction of VR. Engineers often construct and interact with complex three dimensional (3D) models using two dimensional (2D) computer-aided design (CAD) programs.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 13–29, 2018.
https://doi.org/10.1007/978-3-319-91581-4_2

VR offers the potential of a 3D interface to match the 3D models engineers work with, allowing them to interact directly with their work and removing the need to use complex mouse and keyboard controls simply to view and navigate.

Our team conducted several interviews with engineers and managers who have experience developing and reviewing traditional CAD models to determine what the limitations of such technologies are and how they could be made better. We also surveyed engineers for feedback about more specific challenges of using a traditional CAD software and what they would like to see in a 3D VR program. We found that viewing and selecting specific elements of objects, especially those within other objects, proved a challenge to many users; and that getting a sense of scale and layout requires years of experience with 2D software, forcing engineers to create costly 3D prototypes to fully communicate designs to nontechnical users.

Using these insights, we developed a prototype program using the Unity platform to be tested on the Oculus Rift VR device. We tested that prototype with several people with varying levels of experience in engineering. Our major takeaway was that while a VR interface is more intuitive than a 2D one when selecting elements, especially for new users, context is vital: making sure the user understands the environment, the object(s) to be controlled, and the controller itself is necessary for users to get the most out of VR.

The code base for our prototype can be found on Github: https://goo.gl/D8HaLQ.

1.1 Contributions

VR offers those without years of technical experience the ability to explore and understand complex three dimensional objects. This has the potential to revolutionize fields such as engineering, architecture, and healthcare by improving communication between technical experts and non-experts.

Our research shows evidence that VR offers a more intuitive and easier to control interface than 2D CAD programs, and our guidelines will help designers of VR systems create useful and usable applications.

2 Background

2.1 Review of Relevant Literature

Until recently, virtual reality (VR) technology has not been sophisticated enough to emulate many of the tasks that engineers and designers need to complete their work. Two-dimensional (2D) and three-dimensional (3D) CAD modeling software is in no imminent danger of being replaced by VR software, but industry experts agree that VR has potential to replace those systems in the next 10–20 years. As a result, most current VR software is mostly focused on enhancing existing processes rather than replacing them, as users familiarize themselves with the hardware and as more sophisticated software and hardware are developed.

As background for our study, we researched what current user interfaces (UIs) are popular in VR applications, particularly those in an industrial setting. To do so we reviewed white papers, scholarly articles, press releases, marketing materials, industry reports, and popular press articles. We also reviewed videos and presentations from conferences focused on virtual reality, where much of the discussion of UI development takes place. Due to time and resource constraints, we did not review individual patents, instead relying on papers and articles discussing their contents. Our goal is to present the research and work being done on UI design in VR, describe the current state of VR use in industrial settings, identify existing roadblocks to further development and adoption, and explain how those roadblocks are being overcome through innovative UI and technological development.

2.2 User Interface Design

The overarching goal in user interface (UI) design is to accurately and effortlessly translate a users intent into action. In VR, this means using gestures that feel intuitive and natural. Alger explains that in order to make VR technology available and used on a large scale, the barrier to adoption must be very small (Alger 2015a, b, c). Users who are not early adopters do not want to be dissuaded by complex gestures or motions to make the software work; rather, they should be able to instinctively understand how what they see works, as one does with a straw or a hammer. In addition, traditional user experience (UX) elements, such as text, images, graphics, and audio, must seamlessly integrate with whatever system is in place so the user does not feel disoriented.

VR UI designers have identified several potential frameworks for user interaction in a VR environment. Many people naturally reach out to touch VR objects, seeking haptic feedback (Alger 2015a, b, c). Even when they know this is not the way to interact with an object, the instinct is difficult to suppress. Other combinations of wands and cursors used in either hand have been tested but no definitive standard has yet been developed. Another interface that has been tested is a gaze-based interface whereby the user focuses his or her vision on a particular object, and either that focus or a button on the head-mounted device (HMD) performs an action on the object (Samsung 2014).

Another type of UI involves using a virtual keyboard to accept user input (Alger 2015a, b, c), while still others attempt to mimic more recent technology by using swipe, pinch, and scroll gestures (Samsung 2014). These have been developed in part to lessen the difficulty in adjusting to a completely new technology and environment, especially for less technologically-adept users (GDC 2017). Giving feedback, whether haptic, visual, or audible, is necessary to assure users they are performing the intended motion and encourage them to continue (Oculus 2015). Range of motion, vision, and ergonomics must also be taken into account when designing user interfaces, as users will not want to use a technology that is physically uncomfortable.

Designers must also figure out how to adjust 2D objects for a 3D environment. Mike Alger discussed needing to have a 2D screen within his 3D environment so users could, for example, read emails which are designed for 2D reading. At the 2014 Samsung Developer Conference, Alex Chu explained that users found 2D thumbnails easier to understand, even if translating the content into 3D thumbnails was more visually appealing. Depth, usually not an issue in 2D, plays a prominent role in 3D design as users can feel disoriented or even scared if objects do not appear to be correct. At the 2016 Google I/O conference, the Google Daydream team discussed using layers similar to those found in Adobe Photoshop to mimic the parallax effect, a phenomenon human brains use to perceive a change in perspective (Google Developers 2016).

2.3 Current Industrial Use

Virtual Reality. To date, most virtual reality applications in industrial settings have focused on three major areas: employee training, employee safety, and prototype/design review. Every company that has embraced VR for these purposes hopes to takes advantage of remote collaboration capabilities and take advantage of a cheaper, more efficient form of reviewing products and processes before beginning an expensive and time-consuming prototyping process. Most major engineering and manufacturing firms purchase technologies from companies that specialize in developing VR software, such as ESI Group and EON Reality, whose products can be tailored to a specific industry and use case.

An Invensys report from 2010 identified that the most challenging aspect of using VR in process industries is that users must maintain a focus on collaboration and learning while being limited to a single-user experience provided by the hardware, normally a headset (Invensys 2010). Additionally, users need to believe they are actually immersed in a virtual world in order to interact as they would in a real-world setting. However, due to technical limitations, it is neither possible nor desirable to make every object dynamic, so the program must track the state of every object in its memory and render any changes in real-time. Real-time rendering is both the most useful and among the most challenging aspects of VR in an industrial setting, according to experts.

Current software used in industrial settings focuses more on visualization and positioning to solve problems concerning things like safety, ergonomics, and design. For example, Ford uses its VR software to visualize the interior of its automobiles so designers and engineers can see the layout from a drivers perspective before manufacturing the first prototype (Forbes 2014). Ford's VR program has been used to launch 100 prototype vehicles, leading to a 70% reduction in worker safety accidents and a 90% drop in ergonomics complaints (Martinez 2015). Other auto companies, such as Volkswagen and Jaguar Land Rover, have also integrated VR technology into their assembly lines in various capacities.

These renderings are also shared with executives who may offer their own input, but in this case the UI is more limited, allowing for viewing but not editing. Even those programs that allow for editing objects are used only for small changes, such as simple positioning or extruding, and are not used to build an object from scratch (PricewaterhouseCoopers 2016). These limitations are due more to current technological restrictions than practical use of the technology;

Goldman Sachs predicts that CAD and CAM software are the most likely technologies to be disrupted by VR in industrial settings, estimating that 3.2 million users will regularly use the technology by 2025 (Goldman Sachs 2016).

Augmented Reality and Mixed Reality. Compared to VR, augmented reality (AR) and mixed reality (MR) are slightly more mature fields in the industrial sector. AR and MR generally do not have the same UI challenges that VR does simply because the technologies work a different way. In AR applications, users interact with real-world objects with some kind of virtual overlay; for example, using smart glasses to display a product's information or quantity remaining in a warehouse as the user looks at the product, instead of having to find the product's information on a mobile device or notebook (PricewaterhouseCoopers 2016). Other potential uses include allowing remote experts to see what a technician sees and give instruction from far away, or embedding temperature or motion sensors to detect problems in machines before they have reached a critical point and require a shutdown.

MR is a more complex technology that combines elements of virtual and augmented reality. A recent study by Accenture Technology described MR as a next-generation digital experience driven by the real-world presence of intelligent virtual objects, enabling people to interact with these objects within their real world field of view (Accenture Technology 2016). MR hardware typically makes use of three primary technologies: infrared to map physical surroundings, infrared to capture gestures by the user and others, and natural language processing for voice recognition. Machine learning and artificial intelligence algorithms then piece together a virtual world around the user that he or she can interact with. MR can offer some of the same benefits as VR without the UI difficulties. MR applications can be used for remote collaboration, training, or virtual prototypes of physical objects, similar to existing VR technologies. However, it remains to be seen if the benefits of MR, namely the ability to maintain spatial awareness and real-world interaction, outweigh the drawback of not being fully immersed in the virtual world.

2.4 Current Non-Industrial Use

The most popular use for VR on the market right now is entertainment, specifically video games. Creating and manipulating virtual worlds is not a new concept in video games; games such as World of Warcraft immerse the player in a virtual world complete with quests, characters, and objects and allow the player to communicate with other players. The premier user experience has led World of Warcraft to capture over 50% of the massively-multiplayer online role-playing game market for several consecutive years (The Journal of Technology Studies 2014). Video games typically use either the first-person or third-person perspective, with an avatar being used to represent the user as the user progresses in the game. Some VR UI designers have attempted to use avatars in a similar fashion, particularly those working in manufacturing settings, but this limits the functionality of the program (PricewaterhouseCoopers 2016). Several VR hardware devices have emulated traditional video game hardware, namely joysticks and buttons as inputs, because they are typically intuitive for an early adopting crowd.

Two other notable industries where VR and AR are being adopted, and where UI is a notable challenge discussed in literature, are the military and education fields. Militaries around the world are using VR applications to train both air and ground forces for things like ground training, collaboration, and field medicine. VR enables users to simulate battlefield environments and weapons usage without being exposed to live ammunition, and to learn from remote experts while being exposed to realistic situations. Obviously, the specifics of these programs are not widely known, but it is known that British troops stationed in Germany have used VR software in preparation for deployment to Afghanistan (Virtual Reality Society).

In the education field, VR has mostly been used to give students access to experiences they would not normally have, such as a substitute for field trips. The use of AR has been observed more; the Journal of Technology Studies recently studied the use of AR in education and found that AR could have many benefits for students, including increased engagement and comprehension, without the UI concerns that come with using VR (Antonioli et al. 2014). Furthermore, AR allows students to maintain awareness of their space so teachers dont have to spend as much effort supervising them. But while the technology was very beneficial, it required teachers either have a certain level of technological knowledge or receive specific training to use the technology, costing both time and money. From a UI perspective, the work with younger students proved that intuitive gestures and content can be easily understood by almost any audience, but there is much work to be done to determine which gestures are most intuitive for certain actions.

3 Testing Methods

3.1 Interviews

Our team conducted several interviews with other students and professionals who have experience using CAD software. Some of the interview subjects had worked as engineers in large corporations while others had more academic than professional experience. Subjects backgrounds were in a variety of industries, including architecture, food manufacturing, clothing manufacturing, and electrical engineering, but all had used CAD systems in some capacity and had an opinion on the strengths and weaknesses of CAD software. We attempted to get different perspectives from various kinds of users in our interviews, rather than just engineers, to better understand what role a VR system could play throughout the product development lifecycle.

Interview questions were focused on getting a better understanding of the subjects experience in the industry, their experience with designing models using CAD software, and understanding what problems they faced in building such models. We also asked questions focused on specific use-cases for VR, such as current methods of and potential for VR use in collaboration, feedback, and prototyping. Questions were not asked about any particular software, instead focusing on general descriptions of how tasks were completed using 2D modeling software.

Engineers. In total, we interviewed three subjects who are engineers, in addition to the survey of engineers and the managers who have engineering experience (both described in more detail in later sections). Each subject had experience both in a classroom setting and in a professional setting building CAD models. We distilled their feedback down to three major takeaways:

1. Each part is typically owned by one engineer, and it is rare for multiple people to work on a part in tandem. Typically, when working in a group, engineers are not doing live adjustments to a model; if they need to do that they do it separately on physical sketches. In a group setting, one person drives (controls the mouse, keyboard, and computer) while the others watch.
2. It is difficult to look within an object or see cross-sections. Typically the user must select a part and either hide it or make it transparent to see the parts behind it, which may take several steps. This also leaves the user vulnerable to missing certain parts and is more time consuming than engineers would prefer.
3. Making small alterations is difficult due to the precision needed, and sometimes a user can make an alteration (e.g. a small extrusion or angle change) without realizing it. This can have severe consequences if the user doesnt catch it early enough, but often there is no good way to see the error.

Artists and Architects. Interviews from two architects and two 3D animation artists revealed that their key pain points of currently available 3D modeling softwares are slow rendering times, version control, software stability, and portability between software variants.

Other notable feedback included:

1. Accuracy was more important to architects whereas visual aesthetics was more important to artists when modeling. However, once a model has been finalized, keeping the exact same ratios and measurements was also important to artists.
2. Architects and artists found working on 2D screens is not challenging, mostly because of the level of proficiency they have already achieved with 2D softwares. Some even found current softwares to be easy to use when accompanied by certain accessories like 3D mouse, Wacom tablet/pen, etc. However, both groups initially experienced difficulty using the softwares for the first few months.
3. Artists always work in 3D whereas architects mostly work in 2D and use 3D only when showing the model to clients or other stakeholders and to see it at the end of final modifications to verify the feel of the room.

Managers. Overall, five subjects were interviewed. Interviewees had managerial experience in industrial settings, including industries such as food manufacturing, clothing manufacturing, and medical devices. Each had some experience either building CAD models themselves or working on engineering teams to review and give feedback to such models. Here are three main points from the interviews:

1. Collaboration is not a barrier in the design and review process, as the currently employed methods (phone, email, in-person conversations) are effective. Additionally, some issues must be discussed in-person, such as which materials to use or to discuss more detailed issues of scale, which can be difficult to convey in a purely visual medium. Some issues exist, notably giving timely feedback when working on a deadline, but most subjects agreed that was a people issue more than a technology issue and would likely persist even using VR technology.
2. Scale and layout are issues with current software. It can be difficult to estimate scale in a purely visual medium, and a slight miscommunication can lead to problems down the road. Several subjects pointed out the challenges of designing the layout for a factory and the difficulties if spaces and distances are not accurately estimated.
3. Precision is extremely important. Current CAD software operate on a level of about 5/10000ths of an inch, and margins in many manufacturing outfits are extremely low, so the slightest error can cause a company to lose a lot of money.

3.2 Survey

After conducting user interviews, we theorized that element selection would be a pain point that could be addressed by VR and decided to conduct further user research. In addition to the user interviews we conducted with students and professionals, we conducted a survey of engineers and designers. The goal was to understand what specific problems regarding element selection engineers encounter when using 2D CAD models.

The survey was divided into three main sections: general information/background, experience and difficulties faced with element selection, and other information. Users were given ample opportunity to describe specific problems they faced, and several chose to do so. In total we received seven responses to the survey, with four of the users being engineers or managers of engineering teams.

From the survey responses we extracted three key takeaways related to element selection in 2D CAD models:

1. Element selection is very tedious. It can require frequent zooming in/out to get to the right magnification to select the correct element, and the user frequently has to rotate the point-of-view which causes disorientation when working with multiple or embedded objects.

2. The software would be better if it only allowed selection of one type of element at a time. Users felt that a good feature would be a sub-menu or similar interface that allowed them to pre-determine which type of element can be selected.
3. Multiple selection and embedded element selection would be useful features in VR. Both of these are difficult to do in 2D models. Due to the difficulty of rotating and zooming in/out to find the right item to select, users often accidentally select the wrong element or deselect everything, causing the user to start over and waste a lot of time.

4 Hypothesis

Following our interviews and survey, we decided to focus on element selection, as that capability was widely applicable and commonly mentioned as difficult. Element selection affects many different functions of using a model, including scaling, extrusion, movement, and testing.

We hypothesized that VR software would grant its users a clearer sense of scale and better three dimensional precision than 2D software, and thus be more intuitive and easier to control. However, the extra arm movement and risk of disorientation using VR could cause discomfort compared to mouse, keyboard, and monitor.

5 VR Prototype User Testing

We created a VR interface in which users could perform basic element selection in order to test element selection tasks on desktop versus virtual reality. We used the Unity game engine and the VR Tool Kit framework to create an Oculus Rift application in which users could select elements of objects. A video introduction

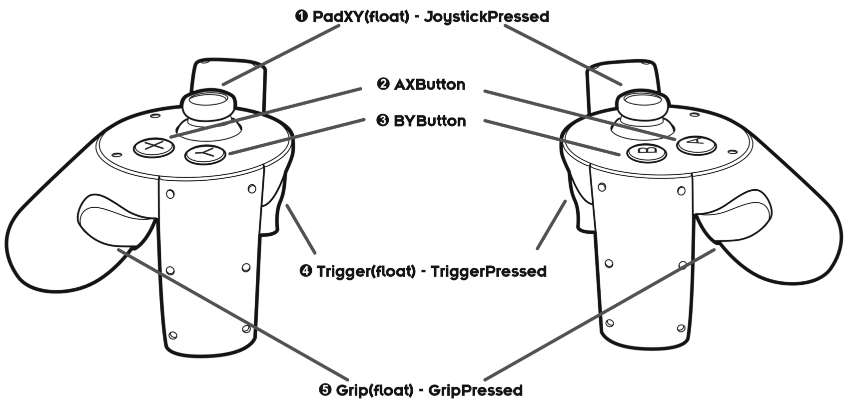


Fig. 1. Our prototype was designed to use the Oculus Rift's motion-tracked controllers as seen here.

of the controls may be viewed at the following link: https://www.youtube.com/watch?v=0fG9dh5uC04 (Fig. 1).

5.1 Prototype Functions

Navigation

- The grip buttons, on the side of the controllers, are used for navigation.
- To move, hold one grip; you can push or pull yourself towards objects.
- To rotate, hold one grip and push the joystick left or right.
- To grow or shrink, hold both grips and bring the controllers towards or away from one another.

Selection

- Hover the controller over a valid element, and that element will light up, indicating it's in focus.
- Press the trigger to select the item in focus, turning it red.
- Hold the A or X button to be able to select multiple objects. (Like holding the control key on a keyboard while choosing files.)
- You can filter what kind of elements you want to select by using the joystick to bring up the *Element Type Filter Menu*.
  - Pushing it in any direction will bring up a radial menu with element types.
  - By pushing the stick towards one type and releasing it, you will only be able to focus on and select elements of that type.
  - The types are (from the top, counterclockwise): Vertex, Edge, Face, Object (not functional at time of testing), All.

Controller Legend

- Press the B or Y button to bring up a legend with instructions for these controls.

5.2 Environment

Selectable Objects. We used two selectable objects in our prototype: an outer blue cube, and an inner green cube. As Unity does not have native support for recognizing and differentiating between vertices, edges, and faces in its 3D objects, we created these cubes out of spheres, cylinders, and thin blocks.

One of our design priorities was that **users should always be able to see what is currently selected and what they are about to select.** We made several design decisions based on this:

- Selected objects are bright red, which stands out against the blue and green unselected states.
- The object that will be selected if the user presses the trigger gains a hover state as represented by increased brightness.
- If there are any objects in between the users head and their controllers, these intervening objects will turn transparent, so the user can always see their controller locations (Fig. 2).

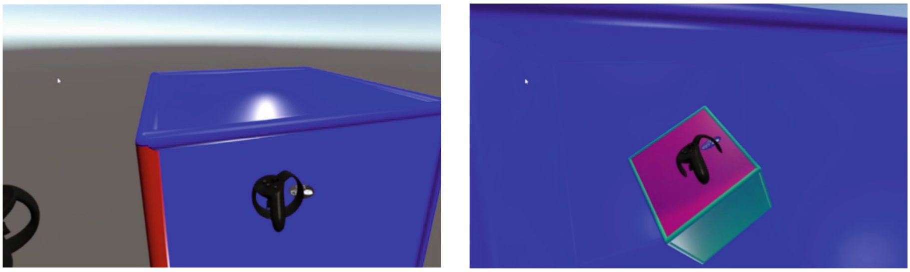


Fig. 2. *Left:* User has selected an edge (red) and is hovering over a face of the outer blue cube, as indicated by the faces increased brightness. *Right:* The outer cube's face turns transparent as the user reaches through it to select a face of the inner green cube. (Color figure online)

Element Type Filter Menu. We designed the element type filter menu so users would have greater control over what types of elements they wanted to select: vertices, edges, faces, objects (not implemented for this prototype), or all types. We reasoned this would be a powerful tool for quickly selecting small elements such as vertices without accidentally selecting larger surrounding elements (Fig. 3).

Fig. 3. *Left:* User can hold the joystick to pull up an element filter menu that will let them select particular types or all types of elements. *Right:* The controller always has a status symbol that indicates what elements it can currently select.

Controller Legend. Finally, we added a controller legend so that users would be able to reference which controls on the likely unfamiliar Oculus Touch controller mapped to which function. Lines go from the control descriptions to the buttons and triggers that activate them (Fig. 4).

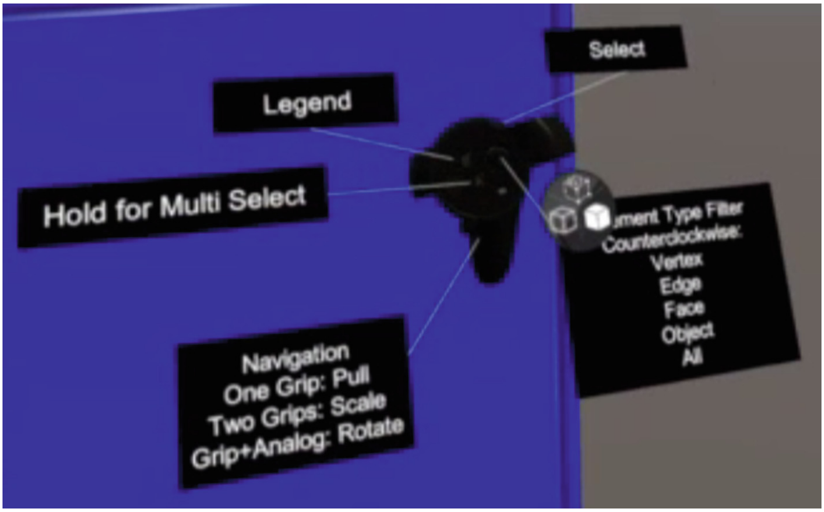


Fig. 4. The controller legend serves as a quick reminder of functionality for the user.

5.3 Testing Procedure

Users were given a brief overview of our project and a quick explanation of the testing procedure. They were given instructions about what to expect and asked to think out loud, whenever possible. We created a 2D model in SolidEdge that was similar to the 3D VR Prototype. Users were given either the 2D or 3D model first, and then the remaining model, in a randomized order.

For each model, users were asked to perform simple tasks using the inputs available (mouse and keyboard for 2D, VR headset and controllers for 3D) and to describe their thinking and emotional state when possible. In the case of the 3D model, users were given 1–2 min for self-exploration before any task was given to get oriented to the 3D environment. Users were given a maximum of 5 min per task, and if they were struggling we offered a hint which users could accept or reject.

After finishing all the tasks on a given model, we asked users to rate their experience on both models using a Likert Scale in four dimensions:

1. Intuitiveness of completing the tasks (1 = not intuitive, 5 = very intuitive)
2. Ease of performing the controls necessary to complete the tasks (1 = not easy, 5 = very easy)
3. Physical discomfort while completing the tasks (1 = no discomfort, 5 = high discomfort)
4. Mental discomfort while completing the tasks (1 = no discomfort, 5 = high discomfort).

Further retrospective probing was conducted on any interactions they found surprising, useful, pleasant, etc.

We realized midway through the interviews that the scales were confusing; high scores were positive for the first two questions, but high scores were negative for the second two questions. Each time we explained specifically what a given score meant, so the users had no confusion about their answers. In the interest of consistency we did not alter our results in any way. All users volunteered for the testing and were neither compensated nor compelled to participate. Each user interview was video recorded, as was the screen on which the user performed the tasks on both models. Each user signed a consent form agreeing to be recorded and confirming they were not compensated nor compelled for participation.

6 Findings

Table 1 shows the results of the questions we asked users following their completion of each model.

Table 1. Feedback from user testing on 2D and 3D models (n = 9)

	2D CAD model	3D VR model
Intuitiveness	1.50	3.67
Ease of control	3.00	4.00
Physical discomfort	1.33	1.67
Mental discomfort	3.78	1.56

Our findings show that on average users found the VR model much more intuitive than the 2D model. This may be biased, as the users we tested the models with generally had less CAD modeling experience than several we interviewed. However, part of our hypothesis was that those with less modeling experience in particular would prefer the intuitiveness of a 3D model. Users also generally found the VR model easier to control, though most users had no trouble using a keyboard and mouse. Most users explained that once they got the hang of the controls in VR, which took about five minutes for most people, they were able to execute the actions very naturally, whereas using a mouse and keyboard they still had to be very precise with their motions, particularly while trying to select individual vertices or edges.

Users did not report much physical discomfort on either model. One user did report some physical discomfort using the VR model, but he explained beforehand that he had used VR headsets before and always experienced some disorientation while using them. Users felt significantly less mental discomfort while using the VR model, and this was evident while watching the users attempt to complete the tasks on both models.

These results confirm our hypotheses that users would find completing tasks easier on the VR model, and that non-technical users would feel more comfortable using the VR model than a traditional CAD model.

7 Analysis and Recommendations

After reviewing our findings, we arrived at three elements that are crucial to a useful and enjoyable VR user experience.

7.1 Tutorial

Given the lack of standardization and general user knowledge of VR user interfaces, a tutorial that acquaints the user with the capacities and intended use of an interface is vital. We expected that it would be obvious that the user needed to touch the cubes directly to select them, but it wasn't: some users looked for a laser pointer or some other way to select, while some thought that they could do it remotely using the element type filter menu. Once we verbally introduced users to the idea of moving themselves over to the cube and touching it, most of them took to it like fish to water; but before that initial instruction, they floundered.

Familiarization with the hardware (i.e. the controllers) is important as well. Some of our testers had never used any kind of game controller before, and later confessed to us that they "did not know all the buttons were functional." If the user doesn't realize a button exists, they won't experiment with it.

The controller legend was useful to most users as a reference, but many found it overwhelming as a first introduction to the controls. We suggest that functions should be taught one at a time, giving the user the opportunity to practice in between each lesson, and that the user have the ability to refer back to previous lessons at will.

7.2 Environmental Context

In 2D CAD programs, it's sufficient for the user to manipulate objects on an otherwise featureless grey background; but in virtual reality, **the relationship of the user to the environment and the objects around them** is incredibly important. Placing our users into a featureless grey plain at best left our users with no frame of reference for their interaction with the cube, and at worst disoriented some of them. Even something as simple as a plane to function as the ground would have helped with this.

The relationship between the user, the environment, and the object(s) in the scene creates further expectations for the user. In our prototype, the cube was the only thing in the environment other than the user; thus, many of our users expected that they would be moving, rotating, and scaling the cube, as in the 2D program. Several users were surprised when they found out they were instead moving themselves. Had we included other selectable objects, or even something like a pedestal for the cube to rest on, we could have corrected for this.

The size of the user as compared to the object(s) in the scene is also important. Different sizes afford different interactions: a user is going to behave differently with something the size of a Rubik's cube than something the size of a shipping container. This is also an ergonomic consideration - if the user has to use large movements to climb all over an object in order to select elements of it, they will quickly get tired. Thus, the tradeoff between large objects which allow precision, and small objects which can be manipulated ergonomically, is one designers should consider carefully.

7.3 Action Feedback

Our users enjoyed getting feedback about their actions. At its most basic, that meant seeing the virtual controllers move when they moved them in physical space. Most of them expressed relief that they could see the small cube simply by reaching or leaning through the big cube in VR, compared to the complex maneuvers necessary to do so in 2D.

We also received mostly positive feedback about the easy to see selected state, and the visual reminder of what mode the Element Type Filter was in prevented a lot of confusion. The haptic and audio feedback from changing modes on the filter was also valuable. There was some frustration when users didn't receive enough visual feedback: the increase in brightness that indicated the hover state wasn't always visible due to ambient lighting conditions.

We believe that visual, audio, and haptic feedback should be available for almost every action the user takes or mode they select. When a user is displaced into virtual reality, they lose feedback that they've lived with all their life such as seeing their hands; failing to introduce alternate forms of feedback in a design will inevitably result in confusion and disorientation.

8 Conclusion

To revisit our hypothesis, our usability testing showed that using virtual reality for 3D modeling addresses two important pain points in existing modeling softwares: steep learning curve and difficulty of control. Our results support this claim, as our VR prototype outperformed SolidEdge in intuitiveness, ease of control, and mental frustration.

There are multiple veins of research that deserve to be explored with regards to 3D modeling in VR: more complex object selection; navigation and control refinement; environmental and scale effects; tutorial design; object selection in AR; and element selection using other types of VR devices such as the HTC Vive or phone-powered headsets like Google Daydream, just to name a few.

Applications of VR to element selection extend far beyond the field of engineering. The easy and intuitive interactions with complex objects that VR enables can be utilized in any field that involves 3D objects, and especially those in which these models must be communicated to people without technical training. Architecture clients will be able to quickly understand complex floor plans, reaching into a model's interiors to select rooms or draw paths. Students will be able to learn subjects such as cell biology like never before, switching between cellular systems on the fly and viewing them from any angle or size. Doctors will be able to communicate clearly to their patients both in-person and remotely, selecting and annotating elements of human models to show what is occurring in their patient's bodies.

While it's still early to say exactly how VR interfaces will replace or supplement traditional desktop interfaces as a tool for 3D modeling and communication, our testers commented that VR seemed to be a great medium for viewing and navigating around 3D objects. As software designers become more skilled at creating VR interfaces and users gain experience, we will no doubt see VR taking on increasing importance in a wide variety of fields. We hope that the principles discussed here will contribute to a future that takes full advantage of what VR has to offer.

References

Mixed Reality Brings Real Benefits to Enterprises: Accenture Technology (2016). www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_19/Accenture-Mixed-Reality-POV-Final.pdf

For US Manufacturing, Virtual Reality is for Real: PwC Network in conjunction with Manufacturing Institute, January 2016. www.pwc.com/us/en/industries/industrial-products/library/augmented-virtual-reality-manufacturing.html

King, L.: Ford, Where Virtual Reality is Already Manufacturing Reality. Forbes, Forbes Magazine, 7 May 2014. www.forbes.com/sites/leoking/2014/05/03/ford-where-virtual-reality-is-already-manufacturing-reality/

Downey, S.: History of the (Virtual) worlds. J. Technol. Stud. **40**(1/2), 54–66 (2014)

Immersive 3D Training Environment: EON Reality. www.eonreality.com/portfolio-items/immersive-3d-training-environment/

Antonioli, M., et al.: Augmented reality applications in education. J. Technol. Stud. **40**(1/2), 96–107 (2014)

Readiness and Value of Virtual Reality for Industrial Machinery: ESI Group. https://virtualreality.esi-group.com/sites/default/files/50/readiness-and-value-virtual-reality-industrial-machinery-2013.pdf

Rovaglio, M., Scheele, T.: The Role of Virtual Reality in the Process Industry. Invensys (2010). software.schneider-electric.com/pdf/white-paper/the-role-of-virtual-reality-in-the-process-industry/

Virtual & Augmented Reality - Understanding the Race for the Next Computing Platform: Goldman Sachs Equity Research, 13 January 2016. www.goldmansachs.com/our-thinking/pages/technology-driving-innovation-folder/virtual-and-augmented-reality/report.pdf

DHL Successfully Tests Augmented Reality Application in Warehouse, 26 January 2015. www.dhl.com/en/press/releases/releases_2015/logistics/dhl_successfully_tests_augmented_reality_application_in_warehouse.html

Volkswagen Pushes Further Into Virtual Reality: Fortune. www.fortune.com/2017/07/05/volkswagen-virtual-reality-app-htc/

Virtual Reality in the Military: Virtual Reality Society, 12 July 2017. www.vrs.org.uk/virtual-reality-military/

Virtual reality combat simulation: Virtual Reality Society, 12 July 2017. www.vrs.org.uk/virtual-reality-military/combat-simulation.html

Martinez, M.: Virtual technology streamlines Ford's manufacturing. Detroit News, 16 July 2015. www.detroitnews.com/story/business/autos/ford/2015/07/16/virtual-manufacturing/30243959/

Alger, M.: Visual Design Methods for Virtual Reality, September 2015 (2015a)
Alger, M.: VR Interface Design Pre-Visualisation Methods. Online video clip. YouTube, 4 October 2015 (2015b)
Alger, M.: VR Interface Design Manifesto. Online video clip. YouTube, 6 January 2015 (2015c)
Samsung Developer Connection: VR Design: Transitioning from a 2D to 3D Design Paradigm. Online video clip. YouTube, 19 November 2014
GDC: Integrating 2D UI with VR Environments. Online video clip. YouTube, 23 January 2017
Oculus: Oculus Connect 2: Navigating New Worlds: Designing UI and UX in VR. Online video clip. YouTube, 11 November 2015
Google Developers: VR Design Process - Google I/O 2016. Online video clip. YouTube, 19 May 2016

Design and Assessment of Two Handling Interaction Techniques for 3D Virtual Objects Using the Myo Armband

Yadira Garnica Bonome, Abel González Mondéjar, Renato Cherullo de Oliveira, Eduardo de Albuquerque, and Alberto Raposo(✉)

Department of Informatics, PUC-Rio, Rio de Janeiro, Brazil
yadygarnica@gmail.com, amondejar@inf.puc-rio.br,
{cherullo,ealbuquerque,abraposo}@tecgraf.puc-rio.br

Abstract. Hand gesture recognition using electromyography signals (EMG) has attracted increased attention due to the rise of cheaper wearable devices that can record accurate EMG data. One of the outstanding devices in this area is the Myo armband, equipped with eight EMG sensors and a nine-axis inertial measurement unit. The use of Myo armband in virtual reality, however, is very limited, because it can only recognize five pre-set gestures. In this work, we do not use these gestures, but the raw data provided by the device in order to measure the force applied to a gesture and to use Myo vibrations as a feedback system, aiming to improve the user experience. We propose two techniques designed to explore the capabilities of the Myo armband as an interaction tool for input and feedback in a VRE. The objective is to evaluate the usability of the Myo as an input and output device for selection and manipulation of 3D objects in virtual reality environments. The proposed techniques were evaluated by conducting user tests with ten users. We analyzed the usefulness, efficiency, effectiveness, learnability and satisfaction of each technique and we conclude that both techniques had high usability grades, demonstrating that Myo armband can be used to perform selection and manipulation task, and it can enrich the experience making it more realistic by using the possibility of measuring the strength applied to the gesture and the vibration feedback system.

Keywords: 3D interaction · Virtual reality · Gesture-based control
Myo armband

1 Introduction

Flexibility and freedom are always desired in virtual reality environments (VREs). Traditional inputs, like mouse or keyboard, hamper the interactions between the user and the virtual environment. To improve the interaction in qualitative terms in a virtual environment, the interaction must be as natural as possible, and because of that, hand gestures have become a popular means of Human-Computer Interaction.

Human gestures can be defined as any meaningful body movement that involves physical movements of different body parts, like fingers and hands, with the aim of

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 30–42, 2018.
https://doi.org/10.1007/978-3-319-91581-4_3

expressing purposeful information or communicating with the environment [1]. The interaction through hand gesture recognition is called gesture control, and it is not only used in virtual reality environments, we can find it in applications such control of robots [2], drones [3], electronics [4], and simple applications like games, slides presentation, music, video or camera. Sign language recognition and gesture control are two major applications for hand gesture recognition technologies [5]. Particularly in virtual reality environments, the gesture recognition and gesture control may resolve some problems like losing the reference of the input controls in the real world.

Hand gestures recognition has two approaches until now [6], the first is using cameras and image processing, and it is called visual gesture recognition. The other is using devices that record the electromyography signals (EMG) of the arm and, with the additional information of an accelerometer and a gyroscope, translate them into gestures [5].

Hand gestures recognition using EMG has attracted increased attention due to the rise of cheaper wearable devices that can record accurate EMG data. One of the outstanding devices in this area is Myo armband (https://www.myo.com), equipped with eight EMG sensors and a nine-axis inertial measurement unit (IMU).

Hand gesture control empowers the developers with tools to offer the user a better experience when it comes to selection and manipulation of the objects in virtual environments. Additionally, with the new wearable devices based in EMG recognition and translation into gestures, we are provided with a new variable: the intensity of the electrical activity produced by muscles involved in a gesture. From now on, in this work, we will refer to it as the "force" applied to the gesture.

The objective of this work is to evaluate the usability of the Myo armband as a device for selection and manipulation of 3D objects in virtual reality environments, aiming to improve the user experience by taking advantage of the ability to calculate the force applied to a gesture and leveraging the Myo's vibrations as a feedback system. For that purpose, we propose two techniques (called Soft-Grab and Hard-Grab techniques) designed to explore the capabilities of the Myo armband as an interaction tool for input and feedback in a VRE.

This paper is structured as follows. The next section presents the basic concepts and summarizes related works. Then, in Sect. 3, we present the methodology of this work, including the tasks and techniques proposed, the definitions of the test environment and performing the user's evaluations. Section 4 shows the results of the tests and present the comparison between the techniques. The last section brings the conclusions of this work.

2 Background

2.1 Electromyography Signal Recognition

Electromyography signal recognition plays an important role in Natural User Interfaces. EMG is a technique for monitoring the electrical activity produced by muscles [7]. There is a variety of wearable EMG devices such as Myo armband, Jawbone, and some types of smartwatches. When muscle cells are electrically or neurologically activated, these devices monitor the electric potential generated by muscle cells in order to analyze the biomechanics of human movement.

Recognition of EMG activity patterns specific of each hand movement allows us to increase the amount of information input into the simulation and to realize a more natural, and hence satisfactory, reproduction of the user's gestures. Fundamentally, a pattern recognition-based system consists of three main steps [8]: (i) signal acquisition, in this case, EMG activity acquired by an array of sensors; (ii) Feature extraction, consisting in the calculation of relevant characteristics from the signals, e.g. mean, energy, waveform length, etc.; and (iii) Feature translation, or classification, to assign the extracted features to the class (gesture) they most probably belong to. Once the gesture attempted by the user of the system is recognized, it can be mapped towards the controlled device.

Working directly with the EMG raw signals has two main difficulties:

- Battery life: Processing the raw data onboard using classifiers that are optimized for power efficiency results in significantly better battery's life than streaming that amount of data through Bluetooth.
- User Experience: Working with EMG signals is hard. Building an application that works for a few people is straightforward, but building something that will work for everyone is entirely another question. The signals produced when different people make similar gestures can be wildly different. Different amounts of hair, fatty tissue and sweat can affect the signals, and this is compounded by the fact that the Myo armband can be worn on either arm and in any orientation.

On the other hand, accessing the raw EMG data allows new uses for the Myo armband like prosthetics control, muscle fatigue and hydration monitoring, sleep state monitoring, and identity authentication based on unique muscle signals.

As we will see, most of the works related to EMG-based gesture recognition do not take advantage of the intensity of the electrical activity like a variable to measure the force applied to the gesture.

2.2 Myo Armband

Myo armband is a wearable device equipped with several sensors that can recognize hand gestures and the movement of the arms, placed just below the elbow. Its platform provides some strong functionality that involves techniques of electromyographic signals processing, gesture recognition, and vibration feedback system.

Myo has eight EMG sensors to capture the electrical impulses generated by arm's muscles. Due to differences in skin tissue and muscle size, each user has to take a calibration step before using the gadget. In addition to the EMG sensors, the Myo also has an IMU, which enables the detection of arm movement. The IMU contains a three-axis gyroscope, three-axis accelerometer, and a three-axis magnetometer.

From this data, and based on machine learning processes, the Myo can recognize the gestures performed by the user.

Besides the EMG signals, Myo armband provides two kinds of data to an application, spatial and gestural data. Spatial data informs the application about the orientation and movement of the user's arm. Gestural data tells the application what the user is doing with their hands. The Myo SDK provides gestural data in the form of some preset poses, which represent configurations of the user's hand. Out of the box, the SDK can recognize five gestures: closed fist, double tap, finger spread, wave left, and wave right.

There are a few drawbacks in the current generation of Myo armband. First, the poses recognized out of the box are limited. Second, using only five gestures to interact with the environment may be considered a user-friendly design that largely reduces the operation complexity. However, the limited number of gestures restricts application development. Finally, the accuracy of gesture recognition is not completely satisfactory, especially in a complex interaction. When a user aims to perform complicated tasks that combine several gestures, the armband is not sensitive enough to detect the quick change of user's gestures.

2.3 Literature Review

Although most of the works about Myo are not related to virtual reality, since its very beginning the creators thought it would be a valuable tool in that field. Some works using Myo in VREs are described below.

Some authors [9] proposed a navigation technique to explore virtual environments, detecting the swing of the arms using the Myo, and translating them into a walk in the virtual environment. Other authors [10] describes a mid-air pointing and clicking interaction technique using Myo, the technique uses enhanced pointer feedback to convey state, a custom pointer acceleration function, and correction and filtering techniques to minimize side-effects when combining EMG and IMU input. Besides, [11] created a virtual prototype for amputee patients to train how to use a myoelectric prosthetic arm, using the Oculus Rift, Microsoft's Kinect and Myo. A couple of immersive games was also developed and is available at Myo Market (https://market.myo.com/category/games).

All reviewed works used Myo's predefined gesture set and they did not use the Myo's vibration system to improve the user experience. It is important to note that all of them focus their efforts in the recognition of the gesture and do not take advantage of the intensity of the electrical activity as a way to measure the force applied to the gesture. Additionally, we did not find any study about usability of Myo for selection/manipulation of 3D objects in VRE.

3 Methodology

In order to evaluate the use of the Myo armband as a device for selection and manipulation of 3D objects in VREs, this work proposes two techniques called Soft-Grab and Hard-Grab. The latter leverages a technique developed to assess the force applied by the user during the closed fist gesture. We describe these techniques below.

3.1 Assessing Gesture Force

Based on the observation that the intensity of the EMG signals detected while the users were doing the closed fist gesture was proportional to the force they were doing, and based on previews works about gesture recognition and EMG data [7, 12], we decided to use that intensity as quantity measure of the applied force.

To calculate this measure of force applied by the user during the closed fist gesture, we used the mean of the eight EMG raw channels of the Myo (Fig. 1) and then the mean of those values in a window of ten samples, starting right after the detection of the gesture.

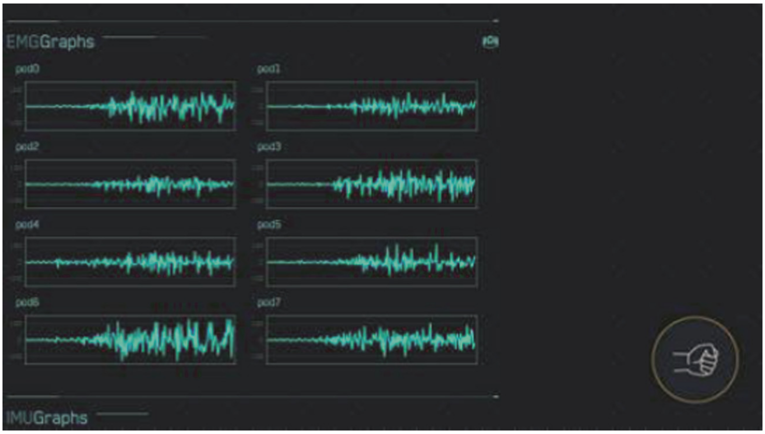


Fig. 1. EMG raw data channels corresponding to the closed fist gesture.

Before users' interaction tests, we made a pilot experiment to determinate a range of force values that we could realistically expect from them during the closed fist gesture. User's EMG signals were measured while they closed their hands using as much force as possible (Fig. 2), and then relaxed the hand again. This experiment was repeated two more times, but instead of closing the hand with force, the users had to squeeze a rubber ball or a hand grip.

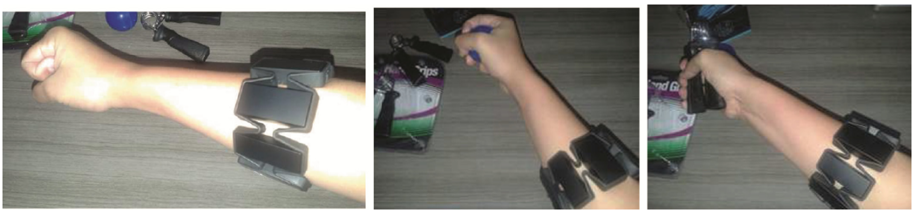


Fig. 2. Doing the closed fist gesture with force, squeezing a rubber ball and a hand grip.

With the obtained data, we extracted the maximum of all the minimum values and the minimum of all maximums values to determinate a possible range of force valid for all users. The obtained ranges and the user comfort of the three experiments were compared and the results were used to setup the tests.

This range of force values was then mapped to a scale of virtual weights that we could use to ascertain if the user was doing enough force during the closed fist gesture to lift a virtual object.

3.2 Proposed Interaction Techniques

Manipulation in virtual environments is frequently complicated and inexact because users have difficulty in keeping the hand motionless in a position without any external help of devices or haptic feedback. Object position and orientation manipulation are among the most fundamental and important interactions between humans and environments in virtual reality applications [13].

Many approaches have been developed to maximize the performance and the usability of 3D manipulation. However, each manipulation metaphor has its limitations. Most of the existing procedures that attempt to solve the problem of selecting and manipulating remote objects, fit into two categories called arm-extension techniques and ray-casting techniques [14]. In an arm-extension technique, the user's virtual arm is made to grow to the desired length so the hand can manipulate the object. Ray-casting techniques make use of a virtual light ray to grab an object, with the ray's direction specified by the user's hand.

The techniques in the present work are based on the ray-casting model. For the scope of these tests, the objects can be moved only in the plane perpendicular to the user's point of view, like in a three-shell game. The virtual environment used in the tests was written in C# using Unity3D 5.6, Microsoft Visual Studio 2015 and the Myo SDK 0.9.0 to connect the Myo.

Usage Scenario. The scenario used to test both manipulation techniques is a big surface with three boxes with different colors: blue, red, and green, and a pointer (Fig. 3).

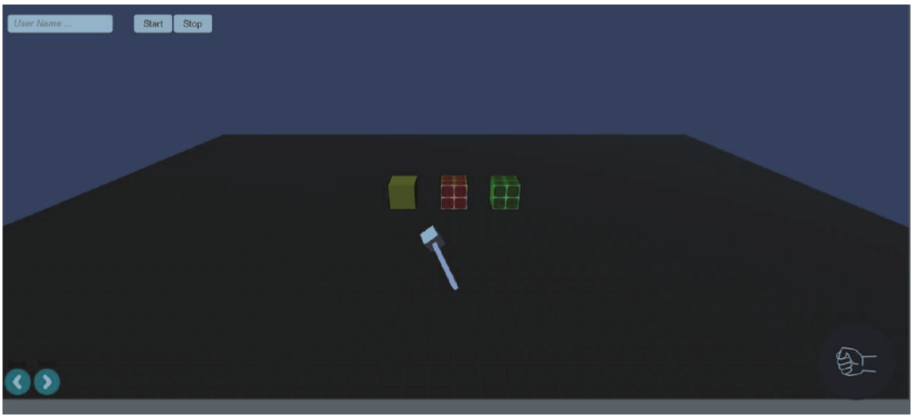


Fig. 3. Test scenario for Soft-Grab technique. (Color figure online)

Pointing was implemented using the Myo IMU and the Raycast method in Physics class from Unity's framework. At the beginning of each test, the user must calibrate the arm's orientation by stretching the arm to the front and making the fingers spread gesture. This calibration procedure can be repeated every time that the user wants to reset the pointer's initial position.

When the ray collides with a box, there is a visual feedback by highlighting the box with yellow color and the Myo's short vibration is activated.

Soft-Grab Technique. While the object is being pointed to, the user can select it by making the closed fist gesture. The Myo's medium vibration is activated when the closed fist gesture is recognized to let the user know that the box is selected and it can be moved.

While selected, the box follows the hand's movements in the (x, y) plane, so the user can position it by pointing the new place for the box. To release it, the user needs to relax the hand and release the closed fist gesture. When the box is released, it is returned to its original color and a large vibration is activated in Myo.

Hard-Grab Technique. In this technique, each box has an associated virtual weight, and as an extra feedback, there is a bar that shows the intensity of the gesture made by the user.

Pointing works the same way, but to select a box, the user must make the closed fist gesture with enough force to offset the virtual weight of the box (as described in Sect. 3.1). When the user reaches the necessary force to lift the box, a Myo short vibration is activated. Manipulation then works the same as in the first technique. To keep the object selected the user must keep the closed fist gesture force within a small range around the activation point. When the force applied falls below this range, the Myo's long vibration is activated and the object is released, returning to its original color.

3.3 User Tests

The user tests conducted in this work followed the guidelines presented by [15] including how user studies should be prepared, executed and analyzed.

Ten participants between 25–35 years old were recruited. Each one voluntarily participated in one test session. Only half of them had experience with 3D interactions and video games, while the other half had no or very little experience. None of them had previous experience with the Myo armband. Two of them were left-handed and the others eight right-handed, so the Myo was used in the preferred hand of the user and that hand was used throughout the test as the hand driving the interaction technique.

The following task list was used throughout the user tests. Before each task, the boxes were restored to their original positions, blue, red, green, from left to right.

1. Explorative (think-aloud) Tasks
   a. Select the left-most box (the blue box), lift it and put it down in the same place.
   b. Select the middle box (the red box), lift it and put it down to the right-hand side of the others.

In the following tasks the user was instructed to not think-aloud, so that an estimation of the time to complete the tasks without interference could be measured.

2. Soft-Grab Test Tasks:
   a. Select the right-most box (the green box) and put it beside the blue box in the left extremity.
   b. Sort the boxes by colors, from left to right, red, blue, green.
   c. Sort the boxes by colors, from left to right, red, green, blue.
   d. Sort the boxes by colors, from left to right, green, red, blue.
3. Hard-Grab Test Tasks:
   a. Select the left-most box (the blue box) and put it beside the green box in the right extremity.
   b. Sort the boxes by colors, from left to right, red, blue, green.
   c. Sort the boxes by weight from left to right, from lightest to heaviest.

The user test sessions were performed in a private room at our university. A laptop was used as the test platform: 8 GB RAM, 1T HD, 2 GB Graphic Card, with Windows 10 operative system. The Myo model MYO-00002-001 was connected to the laptop via Bluetooth. A moderator observed the user's interaction all the time.

The data collected was used to complete two evaluations: first the usability of each individual interaction technique, and secondly the general preference (comparison) of the two interaction techniques.

4 Results

This section describes the results of the user's tests, starting with the achieved usefulness and the measured efficiency for each task included in the test. It follows the qualitative feedback on the perceived effectiveness of each interaction technique. Then, the observed learnability of each technique along with comments from the users are detailed. Lastly, the stated satisfaction of the users on which technique they preferred overall, and for each interaction task, is shown.

The tasks were divided in three groups: Explorative, Soft-Grab Test Tasks, and Hard-Grab Test Tasks. The Explorative tasks were not measured; they were the tasks that the users always did first. The objective of these tasks was to let the user understand how to use the Myo armband. The order of Soft-Grab and Hard-Grab Test Tasks was modified for each user to avoid that the learning of the users influenced in the general result of the test.

4.1 Task Completion: Usefulness

The Soft-Grab tasks group consists, in general, in selecting a box, and positioning it in another place using Myo gesture recognition system. The group has 4 tasks. The first one was to take a box from the right extreme and put it in the opposite extreme, and the other three were sort the three boxes by color, each time with a different arrangement. Table 1 shows the number of users that completed each task.

Table 1. Soft-Grab tasks completion.

	Task completed	Task completed with help	Task failed
Select the green box (right extreme), and put it besides the blue box in the left extreme	8	2	0
Sort the boxes by colors, from left to right, red, blue, green	7	2	1
Sort the boxes by colors, from left to right, red, green, blue	10	0	0
Sort the boxes by colors, from left to right, green, red, blue	10	0	0
Total	35/40 = 87.5%	4/40 = 10%	1/40 = 2.5%

The failure in the second task occurred due to the user misunderstanding the task instructions, this user had very little experience with 3D interactions. The four times when a user needed help to finish a task happened for two users only, and were in the same two tasks. Both were confused about the way on how to close the hand to select the boxes and maintain the hand closed to keep the box selected. In both cases, they expected that once they had selected the box, it would stay selected even if they opened their hands. The moderator had to read again the part of the script that explained how to perform the task. They claimed that they had forgotten that part. From the ten users, just three could not complete all the group of tasks without any help.

The tasks in the Hard-Grab group, in general, involve selecting a box and positioning it in another place using Myo's gesture recognition system and the proposed method to assess the force applied to the gesture. The group is composed of three tasks, the first was to take the box from the left side and put it in the right side, the second was to sort the boxes by colors, and the last was to sort the boxes by weight. It is important to note that in this scenario each box has a virtual weight associated.

In this group of tasks, a single user failed both sorting tasks due to difficulties selecting the boxes (Table 2). The user clearly struggled to reach the minimum force and maintain it to hold the box. The same user completed with help the first task of getting one box from one place and move it to another. To complete the first task, he/she selected the box and could not put it in the indicated place before releasing the box, so the box went down in the middle of the other two boxes. The user then asked if he/she could select it again and put it in the indicated place. The answer was positive and the user completed the task.

Two other users had trouble with the sort by weight task. They wrongly sorted the boxes. Then they selected each box again to measure the weight of the boxes and finally correct the arrangement. They asked if they could correct their arrangements, but since they did explicitly announce that they had finished the task, the moderators allowed them to do it. From the three users that could not complete the task without help or did not complete it, two were right handed and one left handed.

Table 2. Hard-Grab tasks completion.

	Task completed	Task completed with help	Task failed
Select the blue box (left extreme), and put it besides the green box in the right extreme	9	1	0
Sort the boxes by colors, from left to right, red, blue, green	9	0	1
Sort the boxes by weight from left to right, from less heavy to heavier	7	2	1
Total	25/30 = 83.3%	3/30 = 10%	2/30 = 6.6%

In terms of completion, we did not see any important difference between the two techniques and both had good rates of achievement. Also, it was clear that the users who did not complete the tasks were those with less experience in virtual reality environments. To be right-handed or left-handed did not influence the performance and completion rate of the tasks.

4.2 Task Duration: Efficiency of Use

Efficiency was measured by tracking the completion times of the group of tasks: failed tasks were not counted; neither was counted the time when the user stopped to ask for help or to make a question. The tasks were grouped by technique used and the skill required by the user. Table 3 shows the results. We calculated the average and the standard deviation for each group.

Table 3. Tasks execution time.

Technique/Task	Move a box from one place to another	Sort the three boxes by color	Sort the three boxes by weight
Soft-Grab	12.6 ± 3.8 s	25.6 ± 5.7 s	-
Hard-Grab	18.3 ± 2.4 s	29.5 ± 4.9 s	39.8 ± 3.2 s

The first two groups of tasks were completed faster with Soft-Grab technique, but it is important to note that some difference in time execution was expected due to the nature of the techniques. Also, it is important to note that the difference was not so big. Some of the average values of Hard-Grab technique collide with outsider values of Soft-Grab technique, making those hard-grab values to be inside the range of the standard deviation of Soft-Grab technique.

Another interesting point of these measures is that even when the measure of the task "Sort the three boxes by weight" could not be compared with the Soft-Grab technique due the nature of the techniques, its average time and standard deviation shows that its execution time do not differ so much from the task of "Sort the three boxes by color", and the delay time was expected because the user had to select all the boxes to know which one were heavier, and sometimes they need to do that more than once.

4.3 Effectiveness

The effectiveness of the interaction techniques was measured qualitatively by how well the user could use the techniques to select and manipulate virtual objects, based on the think-aloud process, the observed behavior, and the post-test interview results.

Selection and positioning were very similar in both techniques. The difference resides in the force that the user must to apply to the gesture to select the object. Besides that, the key to select an object was pointing it and making a fist gesture, and for positioning the user must point to the new place while the object is selected.

With Soft-Grab technique, the major trend was that the grabbing motion felt natural or comfortable, imitating the way a user would grab an item in real life. Also, they liked that pointing to the box was easy and precise. Two users said that the pointer would be even better if it was shown in the scenario all the time; as implemented, the only visual feedback available is the box highlight when it is hit by the ray cast.

Two users said that, at first, they expected that once they had selected the box, it would remain selected until they had performed the fingers spread gesture. However, both techniques require that they maintain the closed fist gesture to hold the box selected. Relaxing the hand is enough to deselect the box, the fingers spread gesture is not needed for deselection.

Another point that emerged in the interview was the vibration feedback. The users mostly agree and use almost the same explanation for it. They said that at first they could not differ between the three types of vibrations or what they meant, but after two or three times it was very helpful to them to know what was going on.

With the Hard-Grab technique that includes the measure of the force applied to the gesture, the major trend was that it felt very realistic, but it was also more difficult to achieve the selection.

4.4 Learnability and Satisfaction

The major trend for the learnability of both interaction techniques was that the basics for each technique was easy to understand. The major problem was with the Hard-Grab Technique, to learn the force needed to select a given box, and how to maintain the force inside the range required to keep it selected. In the task of sorting by weight, some users were insecure because they needed to lift and put down all the boxes to know their weights, but in general, it takes just a few seconds for them to discover it.

In the post-test questionnaire, we asked the users to evaluate the effort they did to learn how to do each type of interaction (selecting and positioning) with each technique. Soft-Grab required a little less effort for selection than Hard-Grab, but in positioning the difference was much larger, confirming what the users told in the interview.

In general, it was observed a fast learning process; the users were more comfortable with each task, independently of the order of the test applied. The first tasks were always more difficult than the rest of the tests. However, Soft-Grab requires less cognitive effort than Hard-Grab from their point of view.

With the Hard-Grab Technique, the necessity of making more or less force to select a box was in general very well received. The common opinions about it were that it was more difficult, but they could do it and it felt realistic, the users were quite excited about it.

When the moderator asked them which technique they preferred, the majority answered the Hard-Grab Technique. Table 4 shows how many users preferred which technique for each interaction type and overall. The positioning interaction was controversial because some users that preferred the Hard-Grab technique for selection affirmed that they would prefer to simply maintain the closed fist gesture during positioning instead of having to maintain the same force applied for selection. They argued that it is tiresome, and therefore, they actually would prefer a combination of both techniques.

Table 4. Preferred interaction technique.

Interaction/Technique	Soft-Grab	Hard-Grab	Both
Select	2	7	1
Translation/Positioning	6	3	1
Overall	3	7	0

5 Conclusions

This work proposed two selection/manipulation techniques using the Myo's SDK to capture and analyze the spatial and gestural data of the users. Additionally, to take advantage of the new resources that Myo offers, we used the intensity of the electrical activity obtained from the EMG raw data, and we simulated the force that the user was applying to the virtual object. Additionally, we created a feedback system that includes visual and haptic feedback, using the Myo's vibration system.

We evaluated the proposed techniques by conducting user tests with ten users. We analyzed the usefulness, efficiency, effectiveness, learnability and satisfaction of each technique and we conclude that both techniques had high usability grades, demonstrating that Myo armband can be used to perform selection and manipulation tasks, and can enrich the experience making it more realistic by using a measure of the force applied to the gesture and its vibration feedback system. Although, from the interviews, we note that the user's muscle fatigue is an important factor to be deeply analyzed in future studies.

We conclude that the Myo armband has a high grade of usability for selection/manipulation of 3D objects in Virtual Reality Environments. Myo seems to have a promising future as a device for interaction in VRE. More than just navigation, selection and manipulation, it can also be used as a device to input force, offering new ways of interaction in VRE, and in many possible applications like immersive training apps, video games, and motor rehabilitation systems where the possibility of measuring the force applied to the gesture may have a significant meaning. Then, more extensive studies are needed to determine all the advantages and possible uses of the Myo as interaction device in VRE.

References

1. Rautaray, S.S., Kumar, A., Agrawal, A.: Human computer interaction with hand gestures in virtual environment. In: Kundu, M.K., Mitra, S., Mazumdar, D., Pal, S.K. (eds.) PerMIn 2012. LNCS, vol. 7143, pp. 106–113. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27387-2_14
2. Sathiyanarayanan, M., Mulling, T., Nazir, B.: Controlling a robot using a wearable device (MYO). Int. J. Eng. Dev. Res. **3**(3), 1–6 (2015)
3. Nuwer, R.: Armband adds a twitch to gesture control (2013). https://www.newscientist.com/article/dn23210-armband-adds-a-twitch-to-gesture-control/. Accessed 7 Feb 2018
4. Premaratne, P., Nguyen, Q.: Consumer electronics control system based on hand gesture moment invariants. IET Comput. Vis. **1**(1), 35–41 (2007)
5. Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., Yang, J.: A framework for hand gesture recognition based on accelerometer and EMG sensors. IEEE Trans. Syst. Man Cybern.-Part A: Syst. Hum. **41**(6), 1064–1076 (2011)
6. Mann, S.: Intelligent Image Processing. IEEE (2002)
7. Polsley, S.: An analysis of electromyography as an input method for resilient and affordable systems: human-computer interfacing using the body's electrical activity. J. Undergrad. Res. (Summer 2013-Spring 2014), 30–36 (2014)
8. Riillo, F., Quitadamo, L.R., Cavrini, F., Saggio, G., Pinto, C.A., Pastò, N.C., Sbernini, L., Gruppioni, E.: Evaluating the influence of subject-related variables on EMG-based hand gesture classification. In: IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–5. IEEE (2014)
9. McCullough, M., Xu, H., Michelson, J., Jackoski, M., Pease, W., Cobb, W., Kalescky, W., Ladd, J., Williams, B.: Myo arm: swinging to explore a VE. In: Proceedings of the ACM SIGGRAPH Symposium on Applied Perception, pp. 107–113. ACM (2015)
10. Haque, F., Nancel, M., Vogel, D.: Myopoint: pointing and clicking using forearm mounted electromyography and inertial motion sensors. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 3653–3656. ACM (2015)
11. Phelan, I., Arden, M., Garcia, C., Roast, C.: Exploring virtual reality and prosthetic training. In: IEEE Virtual Reality (VR), pp. 353–354. IEEE (2015)
12. Chen, X., Zhang, X., Zhao, Z.Y., Yang, J.H., Lantz, V., Wang, K.Q.: Multiple hand gesture recognition based on surface EMG signal. In: The 1st International Conference on Bioinformatics and Biomedical Engineering, ICBBE, pp. 506–509. IEEE (2007)
13. Nguyen, T.T.H., Duval, T.: Poster: 3-point++: a new technique for 3D manipulation of virtual objects. In: IEEE Symposium on 3D User Interfaces (3DUI), pp. 165–166. IEEE (2013)
14. Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Proceedings of the Symposium on Interactive 3D Graphics, pp. 35–ff. ACM (1997)
15. Rubin, J., Chisnell, D.: Handbook of Usability Testing, Second Edition: How to Plan, Design, and Conduct Effective Tests. Wiley Publishing, Indianapolis (2008)

Surface Prediction for Spatial Augmented Reality

Adam Gomes(✉), Keegan Fernandes, and David Wang

University of Waterloo, Waterloo, ON N2L 3G1, Canada
adgomes@uwaterloo.ca

Abstract. Image projection in spatial augmented reality requires tracking of non-rigid surfaces to be effective. When a surface is moving quickly, simply using the measured deformation of the surface may not be adequate as projectors often suffer from lag and timing delays. This paper uses a novel approach for predicting the motion of a non-rigid surface so images can be projected ahead of time to compensate for any delays. The extended Kalman filter based algorithm is evaluated using an experimental setup where an image is project onto a deformable surface being perturbed by "random" forces. The results are quite positive, showing a visible improvement over using standard projection techniques. Additionally, the error results show that the algorithm can be used in most surface tracking applications.

Keywords: Spatial augmented reality · Virtual reality
Non-rigid surfaces

1 Introduction

In spatial augmented reality (SAR) applications, the projection of images onto non-rigid surfaces can pose many issues. As the surface geometry is not necessarily stationary, standard projection techniques can fail to create a realistic experience for the user due to improper image mapping. For applications where realism is of great importance, this can affect how well a user can perform their intended task. An obvious approach to solving this problem is to track the surface geometry and project warped images onto the measured surface. There have been a number of studies that have investigated tracking and projecting onto non-rigid surfaces [7,10,11]; however, for quickly changing surfaces, there is no mention of how well these techniques perform. When a surface being projected onto is moving quickly, the computational time of processing images, in addition to surface tracking, may cause delays that lead to distortions in the images. To combat this, a prediction scheme can be used to approximate the position of the surface at the time of projection, resulting in a smoother experience for the user. The aim of this paper is to show that a prediction based surface tracking algorithm [3] does in fact improve the realism of SAR by running the algorithm on a real-time experiment. There are many industries that could benefit from

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 43–55, 2018.
https://doi.org/10.1007/978-3-319-91581-4_4

using this kind of technology namely the entertainment and fashion industries, and the field of surgical training [2]. Specifically, simulated surgery using spatial augmented reality is a growing method of surgical training that requires non-rigid surface projection, as the surface (the body) changes shape during the course of the procedure. In fashion, SAR is used to display images on clothing for artistic expression, and non-rigid surface tracking lends itself naturally to this application. It is expected that the methods introduced in this paper will be applicable to each of these fields. This paper does not cover any image warping or projection techniques as it is assumed standard techniques will be used for projection. This paper is organized as follows: Sect. 1 discusses the modelling of the non-rigid surface, Sect. 2 introduces the prediction based surface tracking algorithm, Sect. 3 provides a description of the experimental procedure for real-time application of the algorithm, Sect. 4 presents the results of the experiment and Sect. 5 lists conclusions and future work.

2 System Model

To implement a prediction scheme for surface tracking, a physically accurate deformable model that describes the motion of a surface needs to be developed. A large number of deformable models have been studied in the field of computer graphics, ranging from aesthetically pleasing models to physically accurate models. In this research, mass-spring systems are used to model the dynamics of deformable surfaces due to their simplicity, speed and ease of construction. Mass-spring systems are so popular that they are being used for simulations of deformable bodies in new applications [6]. First developed by Provot [9], the mass-spring model represents a surface by an interconnection of point masses, also called nodes, springs and dampers. As seen in Fig. 1, each point mass is connected to all adjacent nodes with structural springs (or dampers), diagonal nodes with shear springs (or dampers) and nodes that are two steps away with flexion springs (or dampers). Thus, point masses can be connected to anywhere from 3 to 12 other nodes.

The dynamics of the system can be written in the state space form:

$$x[k+1] = f(x[k], u[k]), \tag{1}$$

$$y[k] = Cx[k], \tag{2}$$

where $x[k]$ is the state vector containing the position and velocity information of each node at time-step k, $f(x, u)$ contains the nonlinear dynamics of the system and $u[k]$ is a vector of input forces. The matrix C in Eq. (2) selects only the position states from the state vector to be the output of the model.

To account for errors between the model and the real-life plant, a random process $w[k]$, with covariance Q_k, is added to the state Eq. (1) and a random process $v[k]$, with covariance R_k, is added to the output Eq. (2). The state and output equations now become:

$$x[k+1] = f(x[k], u[k]) + w[k], \tag{3}$$

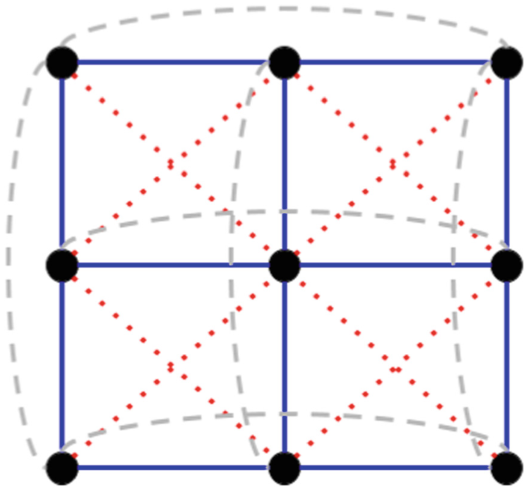


Fig. 1. Connection of point masses with structural springs (blue), shear springs (red dashed), and flexion springs (grey dashed) (Color figure online)

$$y[k] = Cx[k] + v[k]. \tag{4}$$

Although the inner dynamics of the model are linear, the geometry of the model causes nonlinearities (similar to those of a pendulum) that require linearization to be used with the estimation algorithm presented in Sect. 3. Using the standard approach of linearization, the dynamics are converted to the simpler form of

$$x[k+1] = Fx[k] + Bu[k] + w[k], \tag{5}$$

where F is the Jacobian matrix of $f(x, u)$ with respect to x and B is a matrix that selects the inputs related to the velocity states.

With the dynamics of the surface defined in a state space form, the model can easily be implemented into estimation filters; one of which will be used in the algorithm described in the next section.

3 Prediction Algorithm

A common technique to predict states of a nonlinear dynamic system is the extended Kalman filter (EKF) algorithm [1]. The EKF is a extension of the standard Kalman filter, which is an algorithm that uses measured outputs of a system to make estimates of the internal behaviour of the system. The Kalman filter can be used to find state estimates when measurements are corrupted with noise, but can also be used as an algorithm for state prediction. The standard Kalman filter produces the optimal estimate of a system under the condition that the dynamics of system are linear and any measurement or modelling error is Gaussian distributed. The EKF extends the Kalman filter to systems that have nonlinear dynamics. As a result, since the dynamics of the mass-spring system are nonlinear, the EKF can be used to predict the motion of a non-rigid surface. The EKF uses the linearized model, Eq. (5), to update the estimates of the system; thus, it only gives a first-order approximation of system states. As a result, the EKF only gives a "near-optimal" estimate of the system.

Figure 2 shows a simple flow chart of the EKF algorithm where the function $f(x, u)$ describes the dynamics of mass-spring model and the plant is the real-life system on which measurements are made. At each prediction time-step, T_m, the most recent estimate of the non-rigid surface, $x_{k-1|k-1}$, is passed through the mass-spring model, Eq. (1), and a prediction of the surface position and velocity, $x_{k|k-1}$, is made. This prediction is used as the best "guess" of what the surface will look like one time-step into the future. The state covariance matrix $P_{k-1|k-1}$ is sent through the linearized model to produce the predicted state covariance matrix $P_{k|k-1}$. The state covariance matrix gives a description of how correlated the states of the system are to one another at each iteration of the algorithm. This entire step is known as the Kalman prediction step of the EKF. After a new measurement, y, is made from the real-world surface, it is combined with the state prediction $x_{k|k-1}$ and predicted covariance matrix $P_{k|k-1}$ to produce the "near-optimal" state estimate $x_{k|k}$. This part of the algorithm is called the Kalman update step. The state estimate will then be used to create a new prediction for the next time-step, and the algorithm repeats itself. An issue that can arise when measuring the position of a surface is the occlusion of markers. If only measurement data was used to determine the surface geometry, losing vision of a marker would make the projection nearly impossible. However, using this prediction algorithm, the lost marker's position can be approximated using the Kalman prediction, which is a very close estimate of the true position of the marker. This allows occlusion compensation to be nearly free, provided the markers are not covered for an extended period of time.

When running the EKF algorithm for SAR applications, a projector needs to project images on the predicted surface. This can pose issues as the projector takes a certain amount of time to receive and process images from a computer and an additional amount of time to draw a frame. It is well known that projectors suffer from delays when processing images and these delays usually range from 20 ms to 100 ms depending on the type of projector [4]. This delay, T_d, is troublesome when using the EKF for surface prediction in real-time. Since an image needs to be sent to the projector T_d seconds in advance to be projected at the correct time, the EKF needs to predict the geometry of the surface T_d seconds in the future at each Kalman predict step. Now, since measurements are received every T_m seconds, the EKF can only update the state estimate every T_m seconds. An issue arises when the delay time T_d and measurement time T_m do not match (i.e. are vastly different). The time of the current state prediction and the time at which the measurement is made will never be the same. This means the traditional EKF algorithm will not work, as the prediction and measurement times need to line up. To fix this issue, a further prediction, using numerical integration, is made to align the time of the current state prediction with the current measurement. At this stage, a new estimate can be made using the regular EKF algorithm.

When compensating for the delay caused by the drawing of a frame, it is imperative to consider the speed at which the surface is moving compared to the drawing rate. Surfaces that move quickly with respect to the drawing rate of the projector may incur additional image distortion because the projector is still drawing an "old" image. To compensate for the effects of surface movement during the drawing of frames, an inter-frame prediction (IFP) method is proposed. Considering that the update rate of the EKF is T_m seconds, if the cloth's position changes significantly during inter-sample periods, there may be significant error between the prediction and the actual position of the cloth when a new measurement is made. To compensate for this, an interpolation approach is used. As the cloth is moving, the EKF solves for an estimate of the velocity states, and using a first-order approximation, the inter-sample position of every node is calculated. This estimation is based on the assumption that drawing horizontally is instantaneous.

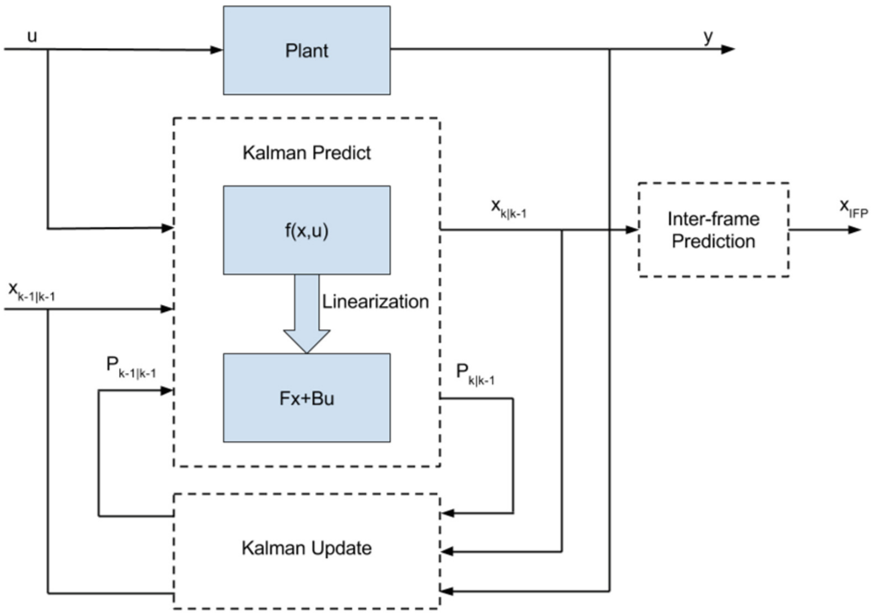


Fig. 2. Block diagram of the EKF algorithm with the mass-spring model

Using the state prediction $x_{k|k-1}$, which was solved with Eq. (1) and the corresponding time-step, $n\Delta T$, where n is the row number and the time-step ΔT is defined by

$$\Delta T = \frac{1}{\text{frame rate} \times (\#\text{rows} - 1)}, \tag{6}$$

the inter-frame prediction can be computed. First the state prediction vector is split into a position prediction vector $p_{k|k-1}$ and a velocity prediction vector $v_{k|k-1}$. The position predictions are then reordered, such that the elements are ordered based on their horizontal position with respect to the projector. More specifically, the first i elements of the position vector would contain the positional information of the first horizontal row of nodes with respect to the projector, the next j elements would contain the positional information of the second horizontal row of nodes with respect to the projector, and so on (Fig. 3).

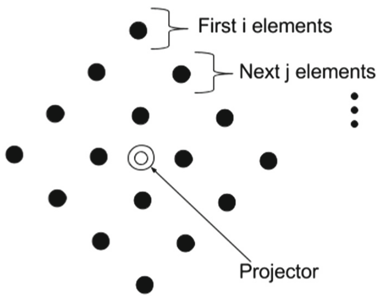


Fig. 3. Orientation of cloth with respect to projector for inter-frame prediction

After reordering the states, the predictions are passed through the state transition function $f(x, u)$, described by Eq. (1). This returns the derivative of the position state predictions, and as a result, the velocities to obtain the next position vector. The velocity vector is then multiplied by a matrix describing the time at which each row of the object is predicted. The result is added to the position estimates to obtain the inter-frame position predictions $p'_{k|k-1}$. At a time t_0, when the system receives a measurement from the cameras, the current prediction at t_0 is combined with the measurement to produce the new estimate. This is done using the aforementioned Kalman update step. Since the time between measurements, T_m, is quite large, the IFP algorithm is run at a time-step of ΔT to counteract the effects of surface motion while drawing. When each new estimate is calculated, every T_m seconds, the Kalman predict step of the EKF is run to create a prediction T_d seconds into the future. This is done to have a prediction of the surface when the projector is ready to draw a frame. This new Kalman prediction replaces the prediction from the IFP algorithm, and the whole sequence repeats itself until termination. The entire EKF-IFP algorithm, compensating for projector delay, is shown in Fig. 4.

4 Experimental Setup

In order to validate the algorithm proposed in Sect. 3, an experimental procedure is designed in this section. The goal of the experiment is to show the effectiveness of using the EKF-IFP algorithm when compared to simply projecting with

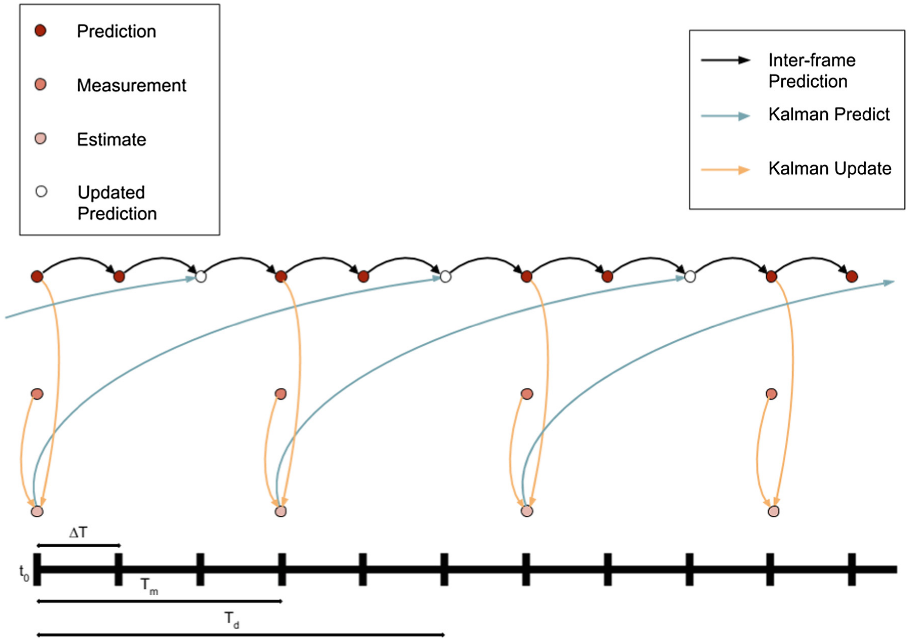


Fig. 4. Timing diagram of EKF-IFP algorithm. ΔT is the IFP time-step, T_m is the measurement time, and T_d is the delay time.

no compensation. This will be done by projecting an image onto a perturbed surface, and using subjective measures to determine whether using the EKF-IFP algorithm is superior to using no compensation. An obvious choice of material to act as the non-rigid surface for the experiment is something cloth-like, as it would be quite deformable. Thus, a towel is chosen as the surface to be projected onto since it is quite sensitive to external forces. A number of different techniques can be used for capturing positional data of the towel's surface, such as image processing techniques or 3D scanning systems; however, for greater data accuracy, a motion capture system is used in this experiment. The NaturalPoint OptiTrack system [8] is an infra-red (IR) camera-based motion capture system that provides positional data, both translational and rotational, within millimeter precision. For this experiment, a three camera configuration is used to measure the position of 12.7 mm diameter infra-red markers. The markers are placed on the towel to match the initial positions of the mass nodes in the model. Specifically, 20 markers are placed on the towel corresponding to a 5×4 node mass-spring system used to model the system. The towel is hung vertically, just as it would be on a standard towel rack, such that all the IR markers are visible to the cameras. An Epson VS240 short-throw projector is placed directly in front of the towel, and below the cameras as to not interfere with the cameras' view. Figure 5 shows the complete experimental setup.

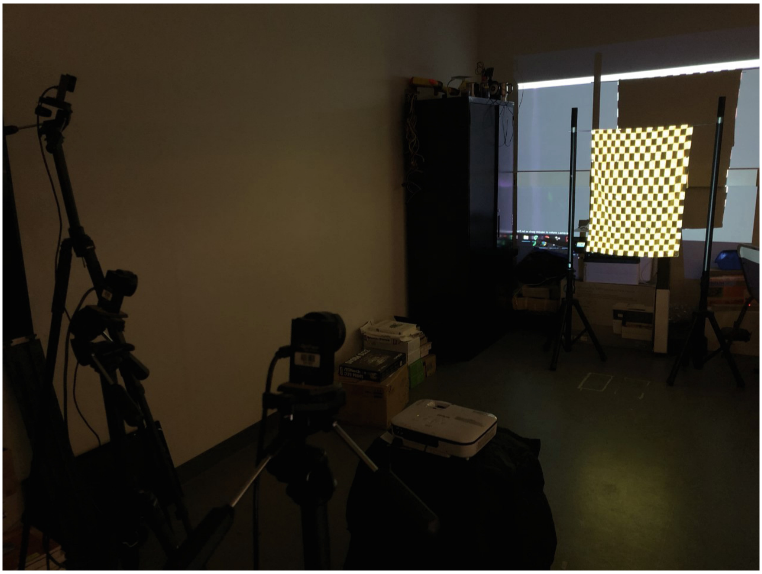


Fig. 5. Photo of experimental setup with three motion capture cameras, a projector and a towel being projected onto.

To implement the algorithm presented in Sect. 3, the mass-spring model parameters need to be chosen so that the deformable model has similar characteristics to the real-life system. Using visual inspection, mass values of 0.025 kg for each node, spring constant values of $300\frac{\mathrm{N}}{\mathrm{m}}$, and damper values of $0.08\frac{\mathrm{N{\cdot}s}}{\mathrm{m}}$ for each spring and damper connection are chosen. It is assumed that any error in parameter choice can be lumped into the process noise term $w[k]$ and will be dealt with by the EKF. The initial position states of the mass-spring model are set to be equal to the position of the IR markers on the towel and the velocity states are set to 0, as the towel is at rest. Since the initial states of the mass-spring model match the initial conditions of the real-life surface, the initial state covariance matrix is set to the zero matrix, as there is no uncertainty between the initial state and the true position of the surface. The measurement noise covariance matrix R_k is set so that the variance of each position state is $0.01\,\mathrm{mm}^2$, and the covariance between any two position states is $0\,\mathrm{mm}^2$ (considered independent). These values of variance are chosen based on the error specifications given by the Optitrack system. The model noise covariance matrix Q_k is chosen to be an identity matrix, as 1 m can easily be assumed to be an extreme upper bound for the uncertainty in node position. Covariance tuning to optimize the EKF will be explored in future revisions of this work.

To begin the experiment, a still image is projected onto the towel when the towel is at rest, as seen in Fig. 5. To project the image on the towel, the system needs to be calibrated so that the computer knows where the projector is relative to the surface. To finish calibration, the timing parameters T_m and T_d

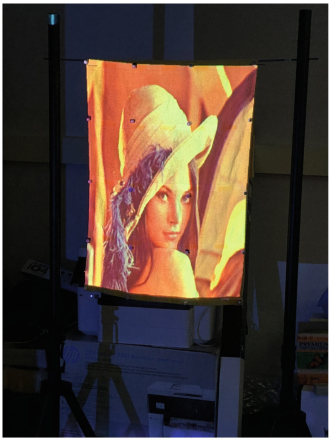
are tuned so that the speed of motion of the model matches that of the towel. The measurement time-step is set to 10 ms and the delay time-step is set to 30 ms. After the system adequately matches the mass-spring model to the towel, a rotating fan is placed behind the towel to create a "random" motion on the surface. This is done to test the robustness of the EKF-IFP algorithm under conditions of randomness. The results of the projection method are visually inspected and predictions of the surface position states are stored to be compared to the real-world values offline.

5 Results

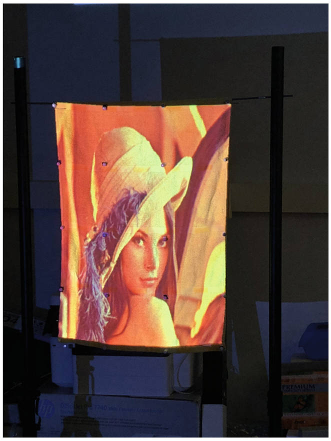
To evaluate the effectiveness of the EKF-IFP algorithm presented in Sect. 3 on the experimental setup described in Sect. 4, qualitative and quantitative methods are used. Qualitatively, the results of the prediction algorithm are visually compared to the results of simply projecting on the surface without any compensation. When the image is projected onto a flat surface (the towel at rest), both projection methods produce the exact same results. However, once the towel is moved, the EKF-IFP method produces more true-to-life results. The images move with the towel, matching its geometry, making it a substantial upgrade over simply just projecting images onto the surface. Projecting directly on the surface clips images and produces generally undesirable results. Figure 6 shows a comparison of a simpler scenario where the towel is put into three orientations. Both projection methods look identical when the towel is at rest. However, when the towel is placed the two other positions (pushed forwards and pulled backwards), the EKF-IFP algorithm produces far more appealing results. Specifically, the uncompensated projection method displays parts of the image past the towel, onto the wall, while the EKF-IFP algorithm "paints" the image on the towel. It should be noted that any distortion in the images when using the EKF-IFP algorithm can be attributed to the short-throw feature of the projector. Short-throw projectors display unequally magnified images so they can be placed closer to walls while keeping the image size intact. Quantitatively, the success of the EKF-IFP algorithm is evaluated using the mean error between the measured position of the markers and the predicted position of the mass nodes. At every measurement time-step, the difference between measured position of node and the predicted position of the node are squared and then averaged. The mean error is defined as

$$E[k] = \frac{1}{N} \sum_{k=1}^{N} \|y[k] - Cx_{k|k-1}\| \tag{7}$$

where N is the total number of nodes (20 in this case), $y[k]$ as defined in Eq. (2) is the output vector, and $x_{k|k-1}$ is the state prediction vector. Figure 7 shows the mean error (ME) between measured and predicted node positions over a 10 second window. It can be seen that after every large input (strong gust from the fan), the ME increases drastically. This is due to the non-anticipatory behaviour of real systems. After this peak in error, the ME exponentially decreases to a

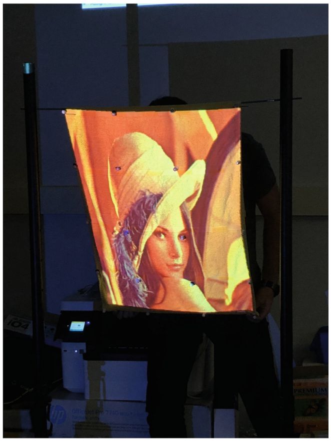


(a) Standard projection: towel at rest



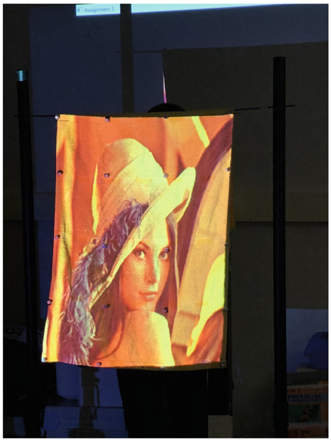
(b) EKF-IFP: towel at rest

(c) Standard projection: pulled backwards



(d) EKF-IFP: pulled backwards

(e) Standard projection: pushed forward



(f) EKF-IFP: pushed forward

Fig. 6. Visual comparison of standard projection and EKF-IFP algorithm

point where there is almost no difference between predictions and measurements. The mean error peaks at roughly 3.5 cm when the towel is most affected by the input force, and 0.6 cm when the towel comes back to rest. This is a very promising result as it shows how effective the EKF-IFP algorithm is at surface tracking. Furthermore, the results imply that the algorithm can in fact be used in any application where surface tracking is required, not just applications of spatial augmented reality.

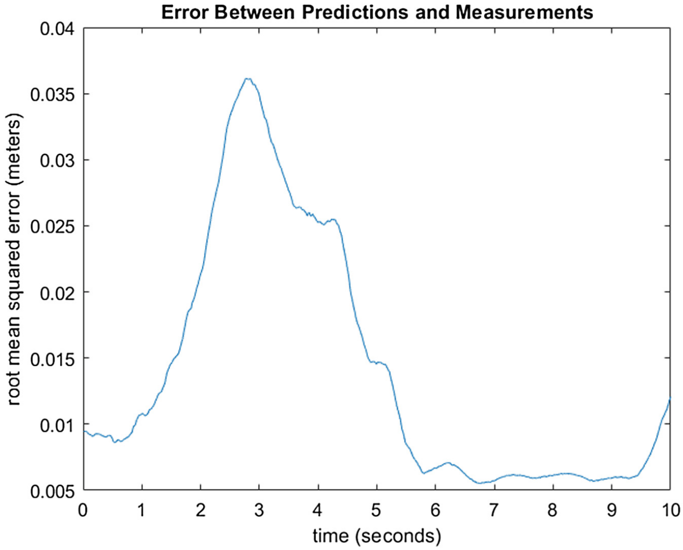


Fig. 7. Mean error graph display the average error between measured and predicted node positions over time.

6 Conclusion

This paper implements a novel technique for predicting the motion of non-rigid surfaces for image projection. The EKF based algorithm, named the EKF-IFP algorithm, predicts the position of a non-rigid surface by using the measured position of the surface while the surface is moving. The algorithm is able to handle the delays often associated with projectors and is robust enough to handle brief occlusions of the surface when measurements are taken. Using a mass-spring system to model the dynamics of a towel, the EKF-IFP algorithm was able to predict the position of the nodes with errors ranging between 3.5 cm and less than 1 cm on average. These results were observed when the non-rigid surface was being perturbed by random forces. If information was known about the input forces, the algorithm would have produced even better results. Using visual observation, the algorithm was also able to project images onto a moving surface with little image distortion. The results show that surface tracking for image projection provides significantly better results than using standard projection techniques in applications of spatial augmented reality. Projection using the EKF-IFP algorithm made interacting with objects far more realistic than using normal image projection, which will make it an indispensable tool for a number of entertainment and training applications.

6.1 Future Work

As the mass, spring and damper parameters for the model were chosen quite arbitrarily, finding parameters that match the surface material properties would allow for more robust prediction. Future work will include using machine learning techniques for parameter identification. Additional future work includes using less obstructive motion capturing systems since the marker based motion capture system is quite expensive and sensitive to environmental conditions. A more cost-effective camera based system, combined with computer vision techniques, can instead be used to capture the position of surfaces in real-time. Although this will likely cause an increase in sensor noise in the system, the prediction algorithm should be able to compensate for the additional measurement error.

A future work of most interest is applying the EKF-IFP algorithm to a scoliosis surgery simulator where a haptic-based robot is used to train surgeons [5]. Current training methods for scoliosis surgery require the use of cadavers. Training on cadavers cannot replicate the "feel" of the surgery and is a very costly approach. A haptic-based simulator is a less expensive alternative that can create a more realistic experience for the trainee. To improve the visual aspect of the simulator, images of the current procedure are projected onto the torso so that the surgeon can interact with the body as they would in the normal procedure. This entails adjusting the visuals to account for deformations of the torso, which is a natural application of the EKF-IFP surface prediction algorithm. The overall goal of this combined system is to replicate both the tactile and visual sensations of the real-life surgery so that surgeons are best equipped to handle this high-risk procedure.

References

1. Anderson, B.D., Moore, J.B.: Optimal filtering. Courier Corporation (2012)
2. Bradley, D., Popa, T., Sheffer, A., Heidrich, W., Boubekeur, T.: Markerless garment capture. In: ACM Transactions on Graphics (TOG), vol. 27, p. 99. ACM (2008)
3. Gomes, A.D.: Prediction for Projection on Time-Varying Surfaces (2016)
4. Livolsi, B.: What it is and why you should care, May 2015. http://www.projectorcentral.com/projector-input-lag.htm
5. Moafimadani, M., Gomes, A., Zabjek, K., Zeller, R., Wang, D.: Haptic training simulator for pedicle screw insertion in scoliosis surgery. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9738, pp. 301–311. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-40244-4_29
6. Nealen, A., Müller, M., Keiser, R., Boxerman, E., Carlson, M.: Physically based deformable models in computer graphics. In: Computer Graphics Forum, vol. 25, pp. 809–836. Wiley Online Library (2006)

7. Piper, B., Ratti, C., Ishii, H.: Illuminating clay: a 3-D tangible interface for landscape analysis. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 355–362. ACM (2002)
8. Point, N.: Optitrack. Natural Point Inc. (2011). http://www.naturalpoint.com/optitrack/. Accessed 22 Feb 2014
9. Provot, X.: Deformation constraints in a mass-spring model to describe rigid cloth behaviour. In: Graphics Interface, pp. 147–147. Canadian Information Processing Society (1995)
10. Punpongsanon, P., Iwai, D., Sato, K.: Projection-based visualization of tangential deformation of nonrigid surface by deformation estimation using infrared texture. Virtual Reality **19**(1), 45–56 (2015)
11. Steimle, J., Jordt, A., Maes, P.: Flexpad: highly flexible bending interactions for projected handheld displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 237–246. ACM (2013)

Real-Time Motion Capture on a Budget

Tami Griffith[1(✉)], Tabitha Dwyer[2], and Jennie Ablanedo[3]

[1] Army Research Laboratory Simulation & Training Technology Center, Orlando, USA
tamara.s.griffith.civ@mail.mil

[2] Cole Engineering Services, Inc., Orlando, USA
tabitha.dwyer@cesicorp.com

[3] Institute for Simulation and Training, University of Central Florida, Orlando, USA
jablaned@ist.ucf.edu

Abstract. The U.S. Army Research Laboratory's Simulation & Training Technology Center, along with Cole Engineering Services, Inc. and the University of Central Florida have set out to leverage commercial technology with the goal of improving realism, and reducing cost for Army training tasks. The focus of this task is to establish a prototype functionality that allows a live person to take control of a virtual character. This is done using the Enhanced Dynamic Geo-Social Environment, which is an Army-owned simulation built upon the Unreal Engine 4.

Commercial games and movies make use of motion capture capabilities to animate characters. This functionality is needed in real-time to allow person-to-person interactions within a simulation. The goal is to have puppeteers that can take over Artificial Intelligence (AI) characters when in-depth interactions need to occur. While AAA games and movie budgets allow for more expensive systems, the goal of this team is to keep the cost well below $10,000.

A market analysis along with this team's experience utilizing and integrating the market capabilities to meet these goals are described in this paper.

Keywords: Virtual training · Avatar puppeteering · Virtual humans Real-time motion capture

1 Introduction

1.1 Background

Motion capture systems are frequently used in the entertainment industry to create an elegant and fluid animation of characters for both movies and games. These systems typically create an accurate and natural representation of a body's dynamic motion. The captured animations are later hand-polished by teams of artists, creating the seamless and natural movement you commonly view within animated films and games. The authors are interested in making use of commercial motion capture capability, including both facial expressions and body movement, to support real-time training experiences. Despite the ubiquitous nature of motion capture systems, it is still unusual for these systems to be used in real-time without an artist to polish the resulting

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 56–70, 2018.
https://doi.org/10.1007/978-3-319-91581-4_5

animations. Because of this unique use case, the hardware and software options have been very limited. A small number of commercial companies at the Game Developers Conference in 2015 promoted real-time motion capture. One of those companies offered a solution that cost roughly $30,000 a year, while another was $1500. The latter was the solution the team pursued for the first prototype of full body and face tracking in real-time. The costs of systems and the equipment needed and can be very expensive to purchase. The authors hoped to make use of low-cost options, though some of most effective real-time body tracking solutions and hardware (such as professionally engineered helmets and cameras) come at a relatively high cost.

This paper describes the research from both a retrospective and future outlook on creating a cost-effective real-time motion capture system, or real-time puppeteering system. Our focus includes integrating real-time motion capture software and hardware into a training game utilizing the Unreal Engine 4.

This paper will also describe analysis of both positive and negative outcomes in creating an overall system. It provides justification for hardware and software solutions and documents how the system evolved through several prototypes. A thorough investigation on affordable real-time body and facial tracking products is described. The initial attempt solely focused on face tracking, without the avatar's body, using a commercial program that is no longer available. The solution evolved to artist-created body gestures/animations, controlled by an Xbox 360 controller. This solution proved unwieldy and felt unnatural for the actor controlling the puppet. Movements were also limited. This resulted in the team exploring ways to improve the prototype. For example, the team experimented with different headgear to keep the camera in front of the actor's face at the appropriate distance and with the right amount of light. A bike helmet and tactical helmet were tried, but there were issues with fitting various users and keeping eyebrows visible to the camera. Most recently, adjustable headgear used for wrestling was used to great effect. This headgear is reasonably priced, and has provided clear view of facial features as well as eyebrows and can easily be adjustable to fit nearly all sized heads. One member of the team modeled hardware and 3D printed it to function as camera mounts on the helmet, and we've used a range of cameras as simple as basic webcams increasing in complexity and cost to $300 depth cameras.

The current prototype uses the HTC Vive with inverse kinematics to resolve body tracking in real-time. We are using a popular, and somewhat affordable, commercial product for face tracking. We also have a working relationship with another facing tracking company that is building a plug-in to support this application in Unreal Engine 4.

Finally, this paper includes a forecast on other solutions that are emerging on the horizon. In particular, Ninja Theory, Ltd is providing good documentation on their motion capture strategy with the game Hellblade: Senua's Sacrifice. The developers of this game attempted to use low-cost capture products, but ultimately ended up using high-end products.

The technology is still in its infancy, and the results have not yet matured into a seamless experience for training. The details of planned future work is included in this paper. This paper is being developed as part of a strategy to grow a community-of-practice into the use of real-time tracking.

1.2 Real-Time Puppeting

The realism of interpersonal interactions in virtual training domains can be an important factor that influences the decisions of a learning population. Some training tasks require very detailed realism, including small-motor movements, eye movements, voice intonation and the ability for avatars to gesticulate while walking in the virtual space. The interaction between automated virtual agents does not yet approach human-to human experiences. This drives our team's interest in increasing the realism of human interactions in virtual environments by using real-time puppeteering [1].

This capability leverages investments across multiple government agencies to optimize costs. The focus is on merging state-of-the-art commercial technologies to support human-dimension training for force effectiveness. The working prototype provides a platform for feedback from end-users and informs the requirements and procurement communities. The authors would like to demonstrate whether this capability is possible at a very low, off-the-shelf cost, with no limits to the number of end-users and no per-seat end-user fees.

There are many potential applications for this technology. For example, Human Resource professionals could use virtual interactions to hone skills in recognizing signs of Post-Traumatic Stress Disorder (PTSD), or indications an individual has experienced sexual harassment or sexual assault. While human-to-human interaction is "the gold standard" to train these skills, the virtual component allows a small number of actors to play various roles to explore racial, gender, sexual orientation, or age discrimination issues. This capability can also be used to breathe life into a virtual landscape; providing the sense that a town is teaming with life. Artificial Intelligence (AI) could control patterns of life within a virtual town. Then as trainees encounter a character, an actor could step into that role and support natural conversation. An entire town could be managed with only a small set of human puppeteers since multiple avatars can be controlled by one person.

2 Stakeholder Goals

2.1 Army Research Laboratory Simulation and Training Technology Center (ARL STTC)

The U.S. Army Research Laboratory Simulation & Training Technology Center (ARL STTC) is focused on research on strategies to improve training for U.S. Army Soldiers. ARL STTC has been exploring using commercial game technology as a way to reduce development time in building simulations to support training tasks. The Enhanced Dynamic Geo-Social Environment (EDGE) is one such simulation platform. EDGE is currently built on the Unreal Engine 4, though it has made use of the Unreal Engine 3 platform in the past. EDGE has been shared across multiple Government agencies with investments from the Department of Homeland Security, the Defense Equal Opportunities Management Institute and the Federal Law Enforcement Training Center, to name a few. The platform is used to explore ways to take emerging technology and apply it to various training tasks. The Army makes use of Computer Generated Forces (CGF) or Semi-Automated Forces (SAF) entities to function as the opposing forces in traditional simulations. Making use of game engines, we are able to make use of Artificial Intelligence (AI). While the AI that comes as part of a commercial game is able to make rudimentary decisions based on the terrain and respond to simple scripted dialog, it is quickly evident that the characters are not driven by a live person. Natural language interpretation is better, but nuances in language still limit verbal interchanges. Until AI is good enough to be used for human-to-human interactions in a simulation, puppeteering can be used to meet the need.

2.2 Impact to the Army

The Army currently supports a great amount of live training events. In some cases, actors are paid to play the role of people in a town. They may even have farm animals such as goats and chickens wandering the town to provide a sense of realism. Soldiers can watch the patterns of life within the town to look for anomalies. Then they can enter the town and speak with passersby, a key leader or law enforcement. As you can imagine, these live events can be costly and logistically taxing. Imagine the cost savings if the town could be living and breathing in a persistent state of a virtual environment when the soldiers log in. A couple of role-players, maybe one female and one male, could hop into a character as a Soldier approaches them, but could be controlled by AI otherwise. Cost savings are important, but it cannot be at the expense of training capability. Our intent, with this research, is to make the virtual environment so that it provides the same training capability as is possible in the live environment to reduce overall training costs to the Army. This represents just one use-case of how this technology can benefit Army training.

3 Prototyping

3.1 Research Problem

The research in real-time motion capture, or puppeteering, initially began to bridge the gap between conversations with a real human and conversations with artificial intelligence (AI) [1]. While great strides are being made in having a natural conversation with an AI character, [2] there is still a large gap in natural response and emotion [3]. Thus, our group began research into real-time motion capture of a human actor puppeting a virtual character, providing the virtual character the natural motion and emotion needed for conversation while at the same time allowing the actor to portray many different characters in a game environment [1].

Budget constraints inspired the goal to reduce end-user costs to achieve a similar real-time motion capture result. Can an effective Real-Time Motion capture solution be achieved on a limited budget? What levels of success can be repurposed for other Government user's? This paper will discuss our path toward finding answers.

3.2 Commercial Game Studio Study

In 2014, in the rudimentary phase of the puppeteering research, the commercial company, Ninja Theory, Ltd, was making similar budget considerations on their new game, Hellblade: Senua's Sacrifice [4]. Ninja Theory, famous for making such games as Heavenly Sword, and DmC: Devil May Cry, decided the company wanted to self-fund their newest game, Hellblade [5]. To successfully accomplish the self-funding goal, Ninja Theory sacrificed several budgetary trade-offs during development [6]. While their proposed budget of just under $10 million is still orders of magnitude larger than our budget, it was beneficial to see the similarity in choices both teams were making [7].

Ninja Theory's decisions in relation to motion capture showed promise for development for the parallel of puppeteering research. Many of the choices made for the puppeteering, including shopping for lights on Amazon, 3D printing camera hardware, using GoPro cameras and mounting hardware, were similar decisions [8]. Ultimately, through trial and error, the puppeteering research, saw some of the same failures, or unacceptable results, that they found with their low budget solutions [1]. Ninja theory, ultimately, borrowed expensive hardware and software from Vicon Motion Capture Products to complete their project [9]. This included high-end motion capture cameras and a professional motion capture helmet, which alone can cost over $3000 [9]. There were many lessons learned through the shared experiences as Ninja Theory explored current technology to meet their goals.

3.3 Software

There were a variety of software components that came together to provide the desired functionality. Code is needed to link the face and body tracking systems to the game engine. The game engine provides the foundational environment and development tools. EDGE is Government owned software built on the UE4 Game engine that allows the researchers to experiment with various technical solutions [10].

Game Engine. The foundation of development is based on the Unreal Engine 4 (UE4) game engine, which many AAA games have been built. The Enhanced Dynamic Geo-Social Environment (EDGE) was built on UE4 and was used for a variety of prototypes. The puppeteering effort is one such prototype, stemming from the concept of filling a gap between human-to-human and human to Artificial Intelligence (AI) conversation [1]. The Unreal Engine 4 was chosen for development for a several reasons:

- The team had significant experience in development in the engine, dating back to prototypes this team developed on Unreal Engine 3. We have been developing in UE4 since 2013 when it was still in beta [1].
- Since our games and research do not generate profit, we are able to use the engine at no cost. One reason we use this game engine is the full access to all source code. The development community is quite large and active, through which we have found many examples and answers to questions through the Unreal Answers

website and the community forums. There is also a marketplace where we have bought 3D model assets and gameplay features providing cost savings over having to develop ourselves.

– The platform UE4 is also a common platform supported by third party commercial developers. An example of this would-be Speed Tree, which builds and ages foliage, plugging directly into the game engine. This is important, because many of the companies currently contain an already developed UE4 plugins or code examples ready to integrate. This is extra work that may be required if we were using a different game engine.

Enhanced Dynamic Geo-Social Environment (EDGE). Virtual training simulations can provide cost efficiencies while also improving training outcomes when used in a blended training concept [10]. Virtual training cannot replace the interaction involved in live training; however, there are opportunities to significantly reduce costs while increasing responder proficiency by applying technologies to support training strategies.

EDGE is a virtual platform developed by the U.S. Army Research Laboratory's Simulation & Training Technology Center (STTC) in partnership with the Training and Doctrine Command (TRADOC), the Department of Homeland Security (DHS) and various other agencies. EDGE is a government owned prototype designed to provide a highly-accurate virtual environment representing the operational environment and utilizing the latest gaming technologies. This collaborative government prototype leverages investment across multiple government efforts to maximize efficiencies (cost savings) by exploiting emerging technology [11].

3.4 Developmental Stages

The puppeteering work has been through many stages of development for over a year and a half. Initially, the goal was simply to accomplish human facial tracking in real-time inside of UE4. Development progressed to adding body animation, and eventually full body tracking. The following section details the stages of development up to the writing of this paper. Earlier work in this area is described in previous papers [1].

The team wanted to answer the question, "Can we convey human emotion in conversation through a virtual avatar?" We thought we could accomplish this through facial tracking software and a camera. We had used software in the past for tradition motion capture and knew it had a real-time component. The team looked toward the commercial market and assessed existing technology. Low-cost options in this space are rare, since they generally target big budget game and movie studios.

Test #1: Facial Capture Only. This effort started with a tracking program that used the Xbox Kinect hardware. The program was developed by a single person and was quite inexpensive. The Xbox Kinect is a depth camera, and we believed that would provide higher fidelity tracking. The integration with UE4 turned out to be labor intensive, and we never made it past the initial integration and test because the tracking software didn't track eye movement. Losing this key feature (eye tracking) removed the sense of realism from the avatar since the eyes were always locked straight ahead. From this task, we learned the importance of tracking eyes.

Test #2: Facial Capture Only, Different Capture Software. The only off-the-shelf options remaining were significantly more expensive, however most companies offered a brief free trial period. Additionally, many companies offered educational or independent video game or Indie (i.e. a game developed without the financial support of a publisher) [12] discount. This allowed for initial testing, and comparison of each of the necessary products. Since the research and prototyping is not for profit, this helped tremendously in accomplishing market research.

The next choice for facial tracking was a larger commercial company discovered at the Game Developer Conference. Their software product also made use of depth cameras for the tracking hardware. The software was tested with the Xbox Kinect. Simultaneously market research was conducted on other depth cameras. There were limited depth cameras on the market, and they were almost five times as expensive as a standard web cam.

While each solution required software development, artwork changes, and hardware creation and/or setup, this capture software required our game characters to have 51 morph target, or shapes, [13] built into the face to control. The art team took existing characters, modified the skeletal mesh to incorporate the shapes, and reimported the character back into UE4. The tasks were coded to an Application Programming Interface (API). The software provided a network stream of shape values for each render frame, and our development team added code to apply those shape values to the character's face in real-time. The camera was placed on a tripod in front of the computer monitor, with the actor sitting and facing the camera.

The initial test with this face tracking solution, as stated, used the Xbox Kinect, but the results were less than ideal. The focal length of the Kinect is very far [14]. This works well for tracking a body, but when focusing on the face, the actor had to sit far away from the desk and camera. This made it challenging to pick up the nuanced movement of the face and eyes. Not knowing if this was a software tracking problem or a hardware (camera) problem, we decided to purchase the depth-camera recommended by the vendor for further testing.

Using the new depth camera, results improved significantly. Exchanging the camera was simple using a dropdown list of cameras within the tracking software. This tracking software required an in-depth calibration for each actor, having them go through a series of expressions and saving that calibration data into a file. It soon became clear that this step was extremely important in future work. After calibration and running with UE4, the results of the expressions and emotion conveyed well through the game avatars.

Test #3: The Desire for Body Language. While face tracking was successful, it was very limiting in scope. For example, the camera was focused on the shoulders and head while the lower body was shown with an idle animation. If the actor expressed anger or sadness, the emotion was betrayed by the body not moving in a sympathetic way. That led to the desire to add body language in concert with the facial tracking.

The original concept did not involve full body real-time motion capture. Rather, initially, the concept was that the actor would function using a game controller to control the body. This mimics other projects and puppeteering throughout history, much like the way Jabba the Hutt's facial features were controlled by radio control in Star Wars: Return of the Jedi [15].

The development effort for this concept was fast paced. The art team created a set of polished character animations for several selected emotions, or postures. The development team mapped those animations to a set of controls on a game controller, and created a user interface that reflected the controlled animations. The design philosophy was to use simple emoticons on-screen and have them mapped to combinations of button presses. For instance, to have the virtual avatar appear nervous, the actor would hold the right trigger and press down on the directional pad of the controller.

The goal was to have this method of acting be intuitive for the actor, while allowing for clean and believable animations. In practice, there were clear benefits and drawbacks to this strategy. Using this strategy, the body gestures could be made to match the facial expressions. However, there was only a finite number of different expressions. The set is nowhere near limitless as one would see in real-life. The biggest issue, though, was that no matter how intuitive the controls were, it was not a natural behavior for the actor. They had to remember that if they were going to say something and appear nervous, they had to press a button combination to get the body into that state. This made it hard to improvise. The actor might respond to a question and quickly look and sound angry on their face, but might forget to change the body using the controller, thus breaking the sense of immersion.

The animations, however, were very clean, having benefited from animator post-processing, unlike raw motion capture data. It was possible to make gestures such as the hands covering the face, which you couldn't do if the actor was covering their face from the camera due to the arms not being collision bodies. This is an issue the team still struggles with today. If arms do not collide they tend to move through one another breaking the sense of realism.

Test #4: Combining Real-time Body with Real-Time Face. Research and development evolved to strive further for a more natural method to capture the body without the controller. As a result, it was decided to find a similar solution of a real-time motion capture system for the body comparable to the one used for the face. To date, the team is unaware of any one solution that handles both, so the focus moved to compile the real-time body capture with the real-time facial capture. Although there has been basic research using the Kinect, as a real-time controller, the research was rudimentary [16], and unfortunately, there were no apparent solutions that would easily integrate into UE4 without significant time and money that would need to be invested.

At the Game Developer Conference additional companies showcased exploring their solutions to the technical field of Real Time Motion Capture. Commercial game companies were now toying with the idea that they could use real-time motion capture to preview their character's performances in the actual game environment [2]. This capability has the potential to save considerable time (and by association money) while shooting motion capture activities [1]. The director can assess the quality of the capture before investing time cleaning up the animation. This capability motivated motion capture hardware and software developers to offer more options for supporting real-time previews [9].

There are two significant drawbacks to most of the vendors providing real-time body motion capture: footprint and price. For many companies, providing a real-time capability does not require additional hardware. A tradition motion capture studio setup involves at least 20 square feet of space, and includes scaffolding and an expensive array of cameras. Our use case called for a much smaller space; basically, a space about 10 feet by 10 feet while the puppeteer stands in front of a computer desk.

Removing motion capture products that required a large footprint left companies that were using inertial motion technology. These are sensor-based systems that don't use cameras to track the body but the position and motion of the body itself [17]. Only one option met the price goals, so it was pursued for further research.

The integration is comparable to face tracking technology. There were a few dozen sensors on the body all sending position, orientation, and velocity information over the network at a frequency matching each render frame. By connecting those sensor points to points on our character's skeleton to drive the in-game position, orientation, and velocity, it resulted in the ability to run both the facial capture server and the body capture server on the same machine. Frame rates remained sufficiently high to see positive results.

Initially the out of the box solution was used to attach the body sensors. A chest mount was used to attach the camera. The chest mount was developed in the laboratory as way to keep the camera pointing at the actor's face. The combination of all the straps, wires, mounts, and batteries made it a bit cumbersome for the actor. The chest mount was not as stable as hoped, and interfered with arm and hand motions (See Fig. 1).

Fig. 1. Chest-Mount prototype with sensor straps

To pursue a more practical use case, the solution moved away from the chest mount and explored head mounted, or helmet solutions. The initial prototype included a low-cost bike helmet with GoPro mounting hardware created within a couple days. To alleviate some of the cumbersome nature of the straps on the body, a motion-capture suit was used. Sensors were attached using hook and loop fastener tape attached to the suit jacket and pants. This simplified preparation for the actor (See Fig. 2).

Fig. 2. Head-mounted prototype with sensors on motion capture suit

The combination of motion-capture suit and camera mounted on the bike helmet became our first successful puppeteering prototype. This is the version we widely demonstrated and documented [1]. The Defense Equal Opportunity Management Institute (DEOMI) saw this prototype and funded additional work to improve the prototype. The team worked with their researchers to develop research studies on learning topics best supported by the technology.

With the new prototype the team began exploring training scenarios that would benefit from this technology. The specific use case involved a group setting with co-facilitators as the trainees. The puppeteer could jump into the body of various group participants, expressing full human emotion in each role. Despite the plan for the characters to stay seated in the group session, the team created strategies to allow puppeted characters to walk using a one-handed controller. The character can move with a walk animation while still being able to control the face, head, arms and hands. This allowed for realistic interactions such as walking down the street with someone gesturing to various points of interest along the way.

Test #5: Test Facial Tracking with New Vendor. During the development process, the company that developed the product we were using for the facial capture was purchased by Apple [18]. Once again, options were limited for a replacement, but the team persisted to make use of a capability provided by a large commercial company known for facial animation in video games with a real-time component.

This facial capture product was a bit different in a few key ways: 1. It makes use of a standard webcam rather than a depth camera; 2. The facial shapes were different, meaning we had to edit all the current character models; 3. Finally, it allowed for quick calibration, meaning it could define the position and outline of the mouth, nose, and eyes. It was not necessary to calibrate each unique face and facial expression.

At first, the results were somewhat less impressive than what we had with the previous vendor. Lip syncing during speech was as fluid as the previous product, but it was more difficult to achieve key emotions that involve smiles, frowns, and furrowed brows. Software updates, such as one that added shape modifiers, have allowed the team to tune a face shape to be prominent. For example, an actor's smile can be magnified to be more obvious on the in-game character.

There were two more issues that needed to be addressed during this phase of research. Low camera frame-rate caused delays in seeing expressions in real-time. Market research indication a paucity of traditional webcams that support 60 frames per second update rate. Thankfully, an eponymous (GoPro-like) camera with high resolution and framerate was available but a video capture card was required for the computer to treat the feed like a webcam feed.

The second issue was the helmet design. The bike helmet was not stable; moving around on the head depending on the size of the actor's head. The helmet was heavy and the additional weight of the camera attached to it pulled the front of the helmet downward. This often caused the helmet to obscure the forehead and brows. Next the team tried using a military tactical helmet. This type of helmet is built to have attachments such as a flashlight connected. Unfortunately, this helmet was even heavier and more cumbersome than the bike helmet. The team settled on using headgear used by wrestlers. This headgear is made up of two solid pieces covering the ears, two straps overhead, and one strap behind the head (See Fig. 3). They are lightweight and can be adjusted to fit any head size. The team's Art Director modeled and 3D printed hardware to mount the camera onto the helmet. This solution, along with lightweight carbon fiber rods, provides exactly the form, fit and function needed.

Test #6: Evaluate New Technology in Real-Time Body Tracking. Throughout this process, the team has experienced elation and frustration. Emerging technology does not always work the way we would like. For example, the body tracking system was frequently plagued with two issues; drift and inaccuracy. Because the motion sensor

Fig. 3. Lightweight adjustable headgear with camera and light assembly

technology is based on magnetic sensors, the sensors drift over time. This means that visual anomalies occur, such as arms starting to clip through parts of the body, or the position of the body in space might shift from center (See Fig. 4). If the actor is sitting on a metal chair, the metal or any wire can create interference causing the virtual body parts to move or be positioned in unnatural ways.

Fig. 4. Character showing some Clipping of Hands

Aside from the issue with drift, there is also an issue with general inaccuracy. Calibration occurs only for specific poses. Even after calibration, arms may not appear in a natural position at the sides of the body. However, the technology does not allow an actor to put the palms of their hands together. Hands are not described as collision bodies, nor do they have "sticky surfaces" that attract the hands together. Rather, while the actor's hands are touching, the character might appear to have their hands a foot apart or crossed over one another, clipping into one another. This is true for both the arms and legs. A good actor could work around these issues and working around the system limitations, but the intent was that movement would be natural so that anyone could play the role of the puppeteer.

4 Outcomes

Though the research area described above is in its infancy, the outcomes show great promise. The resulting prototype has functionality that allows an actor to select a virtual character in a scene and take over that character's control. The actor expresses natural facial expressions and body movements that are played out by the virtual character. The facial tracking system and the body tracking systems are working in conjunction with lower body movement and idle animations. For example, if the actor guides the character forward, using a one-handed controller, the walking animation begins with the lower body, but the actor is able to point and make facial expressions at the same time. This capability goes far in providing a prototype for experimentation.

5 Way Ahead

The prototype is being used to support a wide range of research studies. These will be the topic of future papers. The team continues to do market research into technical advancements and market shifts. Real-time motion capture is a fledgling technology, so no turn-key solutions are currently available. There are many vendors emerging in inertial motion sensors, and this team looks forward to evaluating them to see if there is a solution to the drift and inaccuracy issues described above.

One company has a solution that is showing promise. Their product uses the HTC Vive (See Fig. 5) virtual platform along with inverse kinematics software to achieve real-time motion capture with fewer sensors [19].

The setup involves using the lighthouse (sensors on a stand that work with the HTC Vive) and sensors from the Vive. The Vive sensors track very accurately through the Vive lighthouses. By wearing sensors on both feet, at the waist, on top of the helmet, at the elbows, and while holding the Vive controllers, the entire body can be tracked and moved in real-time (See Fig. 5). Inverse kinematics is software that interpolates the locations of points on the body that cannot be tracked based on the body position of the sensors that are being tracked [19].

The team is building a community of practice on the topic area. The goal is to have this community meet on a yearly basis at the HCII conference to discuss the status of the state-of-the-art and demonstrate progress.

Fig. 5. IKinema's orion full body MoCap on Vive [19]

6 Conclusion

The prototype development described in this paper is intended to build a community of practice interested in building low-cost real-time motion capture capabilities. The development of this prototype, from a game controller feeding specific, clean, animations to a character to the real-time body and facial tracking solution that exists today is useful in informing others with similar interests. As new solutions come to market this team will continue to monitor and evaluate market solutions, providing the results in papers and at conferences as appropriate. The community-of-practice on the topic area is expected to build on this team's research. As this technology area evolves, it is expected that both this team and the community will continue to share status of the state-of-the-art and be able to demonstrate continual progress.

Acknowledgements. The ARL STTC, Cole Engineering Services, Inc., and the Institute for Simulation and Training and the University of Central Florida would like to express our appreciation for the support of the Defense Equal Opportunity Management Institute for resources and expertise provided to this technology area.

References

1. Griffith, T., Dwyer, T., Kinard, C., Flynn, J.R., Kirazian, V.: Research on the use of puppeteering to improve realism in army simulations and training games. In: Lackey, S., Shumaker, R. (eds.) VAMR 2016. LNCS, vol. 9740, pp. 386–396. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39907-2_37
2. Bailey, K.: http://www.techcrunch.com, https://techcrunch.com/2017/02/25/conversational-ai-and-the-road-ahead/. Accessed 25 Feb 2017
3. Merritt, C.: http://www.towardsdatascience.com, https://towardsdatascience.com/a-driverless-world-c3baea61734b. Accessed 27 Sept 2017
4. Summers, N.: The real-time motion capture behind 'Hellblade'. How a tiny team in Cambridge, England, brought Senua to life. https://www.engadget.com/2017/08/08/ninja-theory-hellblade-motion-capture-demo-video/. Accessed 8 Aug 2017
5. Peel, J.: www.pcgamesn.com, https://www.pcgamesn.com/hellblade-senuas-sacrifice/hellblade-unreal-engine-4. Accessed 27 Nov 2017
6. Boyle, E.: Tech Crunch. http://www.techradar.com/news/this-is-the-tech-that-will-make-games-look-better-and-cost-less. Accessed 8 Aug 2017
7. Souppouris, A.: http://www.Engadget.com, https://www.engadget.com/2015/09/16/ninja-theory-hellblade-indie-game. Accessed 16 Sept 2015
8. Hanson, B.: www.gameinformer.com, http://www.gameinformer.com/b/features/archive/2017/04/12/watch-us-push-hellblades-mocap-process-to-its-limits.aspx. Accessed 12 Apr 2017
9. Ninja Theory: Hellblade: Diary 21 - Making a Virtual Human | Real-time performance capture. https://www.youtube.com/watch?v=CmrXK4fNOEo&index=21&list=PLbpkF8TRYizaT6GfMcKBG-RoUOQ6BJRXp. Accessed 17 Mar 2016
10. Maxwell, D.B., Griffith, T.S., Finkelstein, N.M.: Use of virtual worlds in the military services as part of blended learning strategy. In: Handbook of Virtual Environments, pp. 959–1000. CRC Press-Taylor & Francis, Boca Raton (2014)
11. Griffith, T.S., Ablanedo, J., Nenneman, M.: Taking it to the EDGE. In: Proceedings of the International Defense and Homeland Security Simulation Workshop, Bergeggi, Italy (2015)
12. Moore, M.E., Novak, J.: Game Industry Career Guide. Cengage Learning, Delmar (2010)
13. Glanville, S.: Object editor. In: Anim8 or Manual (2006)
14. Brier, D.: E3: New info on Microsoft's Natal – how it works, multiplayer and PC versions. http://old.seattletimes.com/html/technologbrierdudleysblog/2009296568_e3_new_info_on_microsofts_nata.html. Accessed 3 June 2009
15. Marquand, R.: Star Wars Episode VI: Return of the Jedi; George Lucas Commentary (Film). 20th Century Fox (2004)
16. da Silva, D.B., Vidal, C.A., Cavalcante-Neto, J.B., Pessoa, I.N.S., Nunes, R.F.: A robust balance strategy applied to real-time animation data with Kinect sensor. In: 2017 19th Symposium on Virtual and Augmented Reality (SVR), pp. 8–17 (2017)
17. Originating Technology/NASA Contribution: Inertial Motion-Tracking Technology for Virtual 3-D (2005). https://spinoff.nasa.gov/Spinoff2005/ch_10.html
18. Baraniuk, C.: Faceshift: Apple buys Star Wars motion-capture company. BBC News. http://www.bbc.com/news/technology-34920548. Accessed 25 Nov 2015
19. Feltham, J.: http://www.Uploadvr.com, https://uploadvr.com/ikinema-orion/. Accessed 23 Mar 2017
20. McAlone, N.: See the amazing tech behind the original Jabba the Hutt puppet - which managed to fit 3 people inside of it. http://www.businessinsider.com/the-tech-of-the-jabba-the-hutt-puppet-2015–11. Accessed 10 Nov 2015
21. Ninja Theory: Hellblade Development Diary 5: Business of Creation (2014). https://www.youtube.com/watch?v=_f3NMMrQFEE

A Novel Way of Estimating a User's Focus of Attention in a Virtual Environment

Xuanchao He and Zhejun Liu(✉)

College of Arts and Media, Tongji University, Shanghai 201804, China
winggeddreamer@tongji.edu.cn

Abstract. Results from prior experiments suggested that measuring immersion objectively (using eye trackers) can be a very important supplement to subjective tests (with questionnaires). But, traditional eye trackers are not usable together with VR HMDs (Head Mounted Displays) because they cannot "see" an audience's eyes occluded by helmets. The eye trackers compatible with HMDs are not easily accessible to students, researchers and developers in small studios because of the high prices. This paper explores a novel way of estimating a user's focus of attention in a virtual environment. An experiment measuring the relationship between subject's head movement and eyesight was conducted to investigate whether eye movement can be closely approximated by head rotation. The findings suggested that people's eyesight tended to remain in the central area of the HMD when playing a VR game and the HMD orientation data was very close to the eyesight direction. And therefore, this novel way that employs no other equipment than HMDs themselves can hopefully be used to estimate a user's focus of attention in a much more economic and convenient manner.

Keywords: Head movement · Evaluation · VR games · Focus of attention HMD

1 Background of the Experiment

1.1 Introduction

Virtual reality (VR) is a computer-generated scenario that simulates a realistic experience. The immersive environment can be similar to the real world in order to create a lifelike experience grounded in reality or sci-fi [1]. When most people think of virtual reality, or VR, they probably think of hovering cars and time machines as well. While we are still away from traveling through time like Marty McFly, VR is knocking on our door. Technologies like the Samsung Gear VR and Google Cardboard are readily available and affordable for most consumers, as people are trying in more ways than ever to incorporate this innovative technology into their daily lives [2]. Immersive VR systems allow users to experience where they are, whom they are with, and what they are doing as if it was a real experience. In this context, the concept of presence refers to a phenomenon where users act and feel as if they are "really there" in a virtual world created by computer displays [3–5].

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 71–81, 2018.
https://doi.org/10.1007/978-3-319-91581-4_6

Many people think that the VR technology is an emerging technology. In fact, that was not the case. Since 1960s, research institutions have begun to simulate dynamic shapes and sounds using computers, which is followed by the budding of VR. In the next ten to twenty years, early concepts and theories of VR were formed [6]. However, the new wave of VR came in 2016. In the decades since Sega's initial experiments, VR technology has come a long way, and has made significant progress on its early shortcomings [7]. It all comes down to the development of VR headsets. The head-mounted display is the core component of the VR experience. The uniquely immersive experience of VR relies on the head-mounted display. A head-mounted display shows the computer graphics (CG) stereoscopically in front of users' eyes. It is more like a virtual camera through whose lens a user can see a 360-degree virtual environment. It provides wide angle displays covering the normal range of a human being's field of view: at present, Oculus Rift DK2 has field of view of 100° and Oculus Rift CV1 even has field of view of 110°, close to the that of human eyes which is about 120°. This feature makes people forget that they are immersed in the virtual world with head-mounted devices, which basically also allows you to interact with the virtual world with natural human vision. Today, consumers and developers can choose from three basic variations on the theme of stereo-3-D headsets for VR: systems with a dedicated internal screen; headgear that holds a smartphone as the source of the VR input; and augmented reality, in which the headset superimposes 3-D images and data on the user's view of the real world [7].

Although the second type, smartphone-based headset, is much cheaper, the user experience is not as good as the first type which is symbolized by Oculus Rift. In a headset like Oculus Rift, internal optics focus the user's vision onto a screen that is typically only a few centimeters away from his/her eyes. Sensors detect the position and orientation of the head, while the headset uses that information to calculate the images displayed on the screens. More importantly, although the processors, sensors and small high-resolution screens used in dedicated headsets are of the same types used in smartphones, in dedicated headsets they are all optimized for VR [7]. The dedicated headsets like Oculus Rift are connected to high-performance computers or game consoles that generate stereo pairs of images, generally at a rate of 90 times per second.

Meanwhile, because of the revival of VR, more and more game companies start to develop VR games for PC, PlayStation, Xbox One, smart phone or other platforms. There are 2651 VR games for HTC Vive and 1646 for Oculus Rift found on STEAM, one of the biggest online game store [8]. The number of PlayStation VR games is approximately 343 according to Google and Wikipedia [9]. In addition, the advent of VR-ready game engines has greatly simplified the production, not only the CG content creation but also the programming, of VR game, greatly reducing the development costs. As a result, most people would agree that there's never been a better time for VR games than now.

1.2 Related Works

Non-VR digital games have been enjoyed by millions of people around the world for a pretty long time. Modern games often have huge virtual environments for people to explore. Controls are more sophisticated, allowing people to carry out a wider variety of maneuvers in a game. Through the use of the internet people can even play against opponents thousands of miles away. Charlene Jennett mentioned that the success of a computer game depends on many factors. Despite the differences in game design and appearance, successful computer games all have one important element in common: they have the ability to draw people in [10]. Concerning VR gaming, players are more than gamers, they are looking forward to the fulfilled potential of VR and the feeling of "losing" themselves in the VR game world. Such experience is referred to as "immersion", a term often used by gamers and reviewers. Immersion is often viewed as critical to game enjoyment, and is usually the outcome of a good gaming experience. When measuring and defining the experience of immersion in games, Charlene Jennett described it as "the psychology of sub-optimal experience, which clearly has links to the notion of flow (flow is described as the process of optimal experience, the state in which individuals are so involved in an activity that nothing else seems to matter [11]) and CA (Cognitive Absorption, as a state of deep involvement with software [12])". He also mentioned that immersion rather is the prosaic experience of engaging with a videogame [10].

The outcome of immersion may be divorced from the actual outcome of the game: people do not always play games because they want to get immersed, it is just something that happens. It does seem though from previous work that immersion is key to a good gaming experience [10]. The team explored immersion further by investigating whether immersion can be defined quantitatively. In one of the experiments, researchers investigated whether there were changes in participants' eye movements during an immersive task. They use an eye tracker to record participants' eye movements while they were engaged with a task/game. Overall the findings suggested that measuring immersion objectively (using eye trackers) can be a very important supplement to subjective tests (with questionnaires) [10].

1.3 Limitations of Eye Tracker in HMD

Based on these findings, we deduce that immersion of VR games can also be measured objectively (such as task completion time, eye movements). Using questionnaires only is not a precise and reliable enough way to measure immersion of VR games. But, the traditional eye trackers are not usable together with VR HMDs because they need to "see" an audience's eyes, but they are occluded by helmets (Fig. 1).

Eye tracking for HMDs is a natural next step and gained much attention in the research and development sector (e.g., FOVE Inc., Arrington Research, ASL Eye-Track, SR Research, or Sensor Motoric Instruments (SMI)). Even though first attempts started in the year 2000 [13], current inside-helmet eye tracker prototypes are still far from being consumer-ready: the SMI's eye tracker in the Oculus Rift is priced up to USD 15,000, and as for Tobii Pro, USD 28,800 approximately. Obviously, they are not easily accessible to students, researchers and developers in small studios. Moreover, with the ever-changing hardware technology, head-mounted devices may upgrade very quickly. Once installed into an HMD, the expensive eye trackers cannot be detached from it and reinstalled into a new one, which make it more unaffordable and uneconomic for those with a tight budget.

Fig. 1. The typical scenario of using a traditional eye tracker

2 Hypotheses

According to my personal experience, when playing a VR game with an HMD headset on, the eyeball rotation is limited. It is largely complemented by head rotation, allowing the player to look into different directions in the virtual environment. Then, several VR game players were closely observed, and it was found that their heads and bodies almost always moved and rotated when they want to look into different directions in the virtual environment. Such phenomenon can be seen more clearly in a demonstration video offered by SMI, a provider of inside-helmet eye tracking solutions.

In Fig. 2, there are two pictures are capture from a YouTube vide uploaded by SMI (https://www.youtube.com/watch?v=Qq09BTmjzRs). It's not difficult to conclude from the video that in most time the user's focus of attention, visualized as white circles, is near the center of the screen.

Because of this, a hypothesis was brought forward that people's eyesight tends to remain in the central area of HMD screens when they play a VR game. They prefer to move their heads to look at what they want to see, instead of moving eyeballs only. In

Fig. 2. Foveated rendering at 250 Hz from SMI

this way, eye movement can be closely approximated by the orientation of the head, which also means that the data from HMD rotation sensors can be recorded and analyzed to evaluate peoples' focus of attention. Since it can be further used as a low-cost data source for the evaluation of VR games, this approach, which employs no other equipment than HMDs themselves, is hopefully much more economic and convenient for most developers.

3 Method

Though where the subject's eyes were looking at could not be known while wearing a HMD headset, the purpose of the experiment was to evaluate the relationship between subject's head movement and his/her focus of attention by comparing the subject's eyesight direction with the orientation of the HMD. The orientation of the HMD was easy to measure, because it was directly sent by the device as rotation values to the game engine. However, the direction of eyesight was relatively difficult to measure because where the subject's eyes were looking at could not be known directly without an eye tracker. Therefore, a game was designed with the purpose to effectively guide players' vision.

The research team developed an original VR game called "Clock" with the Unity engine, and its functions were as follow. After the experiment began, an alarm clock appeared at a random location in the VR world and the player would immediately hear the ringing it. What the players needed to do was to find out where the clock was and to read out aloud the time shown on its surface. A very simple background was used to ensure no distraction so that the possibility of the player's being attracted to the background was greatly reduced. If the clock stayed in the view for more than 5 s, which meant that the player had found it, it would disappear. Then, the second clock would be generated in a new random position and the procedures described above repeated. There were 5 clocks in total, and the players were encouraged to find them and read out the time on the clock surface as fast as possible. The purpose was to make sure that the players looked at the clock carefully so that the position of the clock could be deemed as the focus of attention.

The angle between two eyesight rays, one going through the center of the HMD view, obtained from the HMD's orientation data and the other going through the center of the clock, was recorded from the beginning, when the first clock appeared, to the end of the test, when all 5 clocks were found, at a rate of 5 Hz. This angle data was saved into a text file. Big angle numbers meant that the alarm clock was far from the center of the screen while small numbers meant that the HMD was pointed accurately at the clock (Fig. 3).

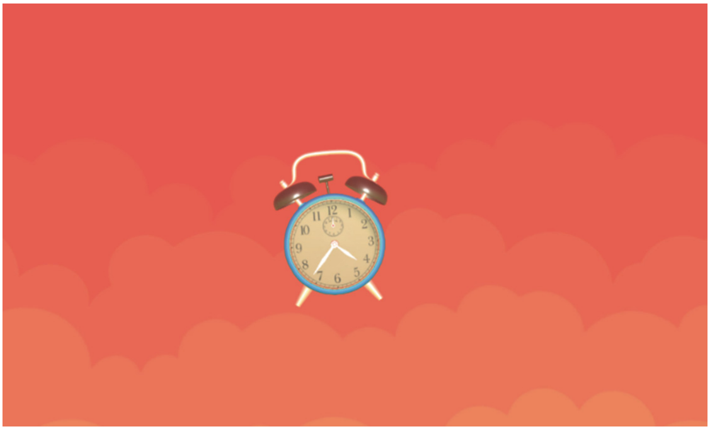


Fig. 3. Screenshot of the game "Clock"

3.1 Participants

20 participants took part in the experiment, all of whom were recruited from Tongji University. The average age was 22.85 (SD = 2.412), ranging from 20 to 26. Ten were male and ten were female. All the participants were willing and healthy. Since this game was intuitive and easy to play, prior experience of VR gaming was not concerned.

3.2 Equipment

Oculus Rift CV1 was used in the experiment as the HMD device. It was connected to a DELL T5810 workstation with an Intel Xeon E5-1660 CPU and an NVIDIA GeForce GTX 1080 graphic card. The high-performance computer was completely capable of running the "Clock" program at a framerate over 120 fps, guaranteeing that there was no bias caused by performance limitations.

The Oculus Rift CV1 had a built-in headset, so no external speakers were used. The "Audio Listener" in the Unity engine received inputs from every "Audio Source" in the scene and played sounds through the headset. For most applications it makes the most sense to attach the listener to the "Main Camera" [15], where the player was located. In this way, the engine could dynamically calculate the relationship between the audio sources and the player, at the same position of the "Audio Listener", and dynamically generated the sound effects correctly. As a result, the player could judge the location of the clock by listening to its ringing via the headset.

3.3 Procedure

Participants took part in the experiment one at a time, and the total duration of one session was about 5–10 min, depending on how much time a participant spent in finding clocks.

An experimenter firstly explained the rules of the game:

1. What a player needed to do was to find the clock and read out aloud the time shown on its surface;
2. Each clock would disappear after having been found for more than 5 s;
3. There were 5 clocks in total, and the player was encouraged to find all of them and read the time on the clock as quickly as possible.

Then, when the participant fully understood the rules, the experimenter helped the participant to put on the HMD device and started the game (Fig. 4).

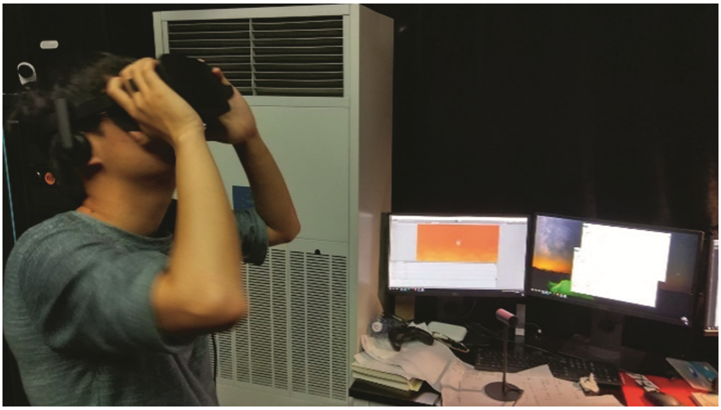


Fig. 4. Participant trying to find the clock

The game had no time limit: it would not end until the participant found all the 5 clocks.

Finally, when the game ended each participant was asked whether the clock remained at the central of the screen after he/she found the clock. The subject answer to this question served as an auxiliary confirmation to ensure that the participant was indeed looking at the clock.

3.4 Results

19 out of all the 20 participants replied "yes" to the question mentioned in the last paragraph of 3.3. There was only one exception: one of the player spent quite some time searching for the clock with no success and began to feel boring, so after he found it, he did not keep focusing on it for long enough. Considering the game design and the players' answers, the ray from the virtual camera to the clock could be thought of as the direction of the eyesight.

The participants' performances were shown in Figs. 5 and 6. The X axis stands for time and the Y axis stands for the angle between the eyesight and the central line of the HMD. The charts are not arranged according to the order of experiments because the subjects are independent to each other and the order was not relevant.

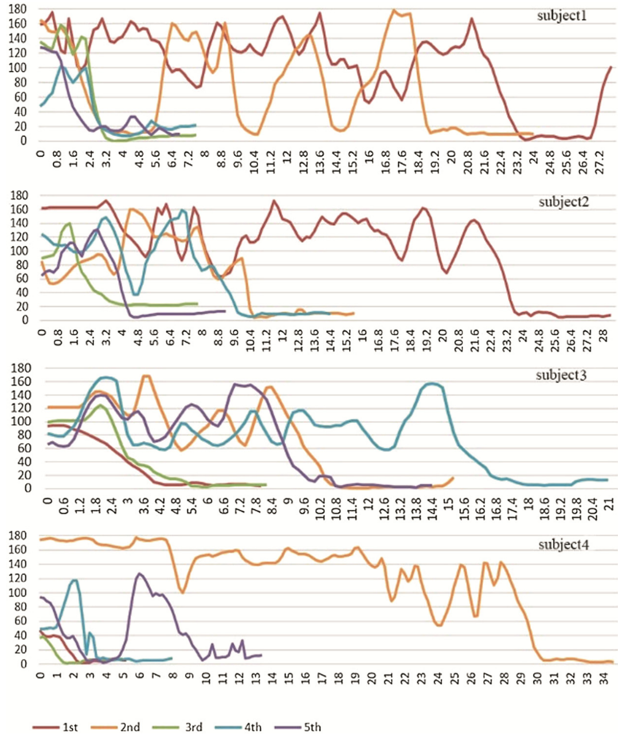


Fig. 5. Player's performance chart (part 1)

Every participant found the clock for five times, so there are 5 lines in each chart, each standing for one search-and-find processes. The left wavy parts of the curves indicate the period of time when the subject rotated his head searching for the clock.

One participant might spend different amount of time searching for each clock, so the lengths of the wavy parts are not exactly the same. For instance, subject 4 spent less than 3 s in finding the clock for the first, third, fourth time. But the second attempt took him as long as 30 s. No correlation was found between the total time used to find a clock and which attempt in order it was. The assumption that one subject could make progress through practice and thus shorten the searching time was not supported.

For the right parts of the curves, the Y value stabilized and approached to 0 for a period of time, which means that the subject found the clock and kept looking at it to read the time on its surface.

The parts whose Y value stay below 30° for more than 4 s are considered as "stable parts", while the rest are considered as the "wavy parts". As time went by the position of the clock on the screen became increasingly steady and close to the center.

Figure 7 shows the situation clearly. Since the alarm clock itself occupies a certain span of angles, we define it as "close to the center" when the angle difference between the eyesight and the HMD central line is below 20°, and "very close to the center" when below 10°. Among all the 100 polylines in the charts in Figs. 5 and 6, Y values of the

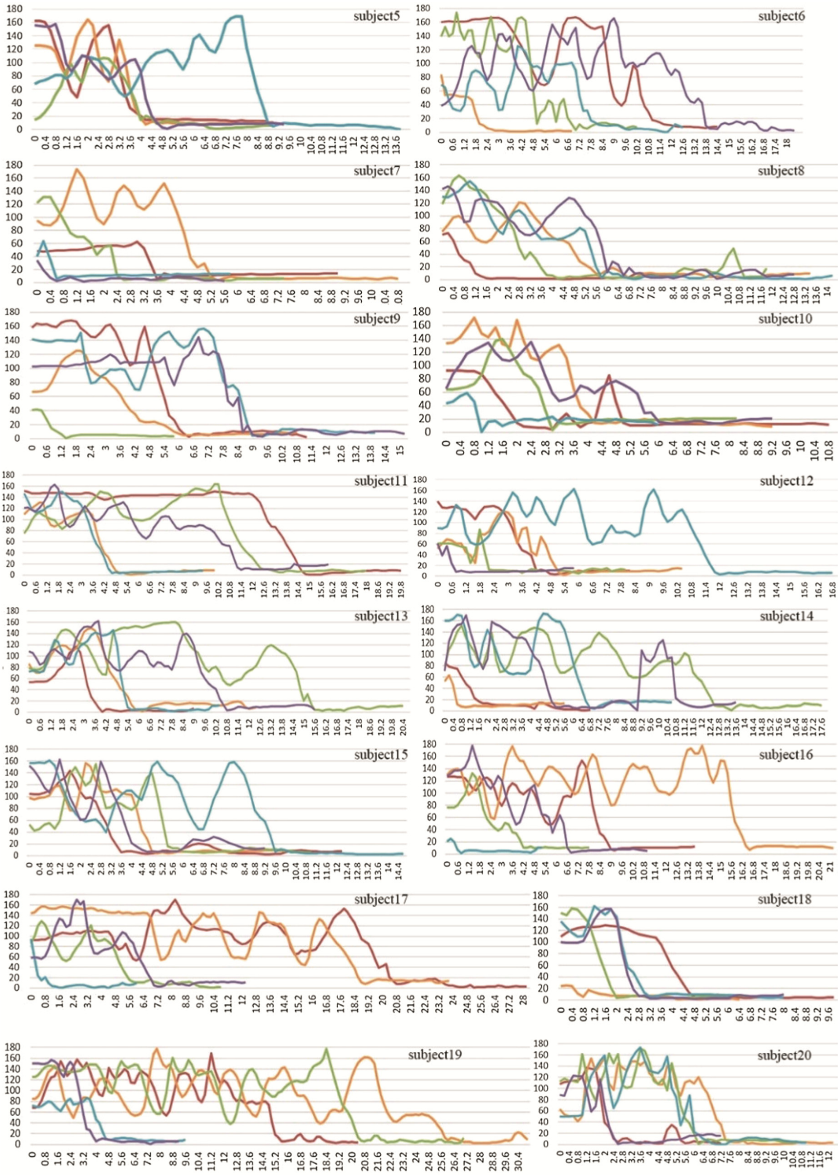


Fig. 6. Player's performance chart (part 2)

stable parts were "close to the center" (<20°) in 88 out of 100 cases (88%), and were "very close the center" (<10°) in 65 out of 100 cases (65%).

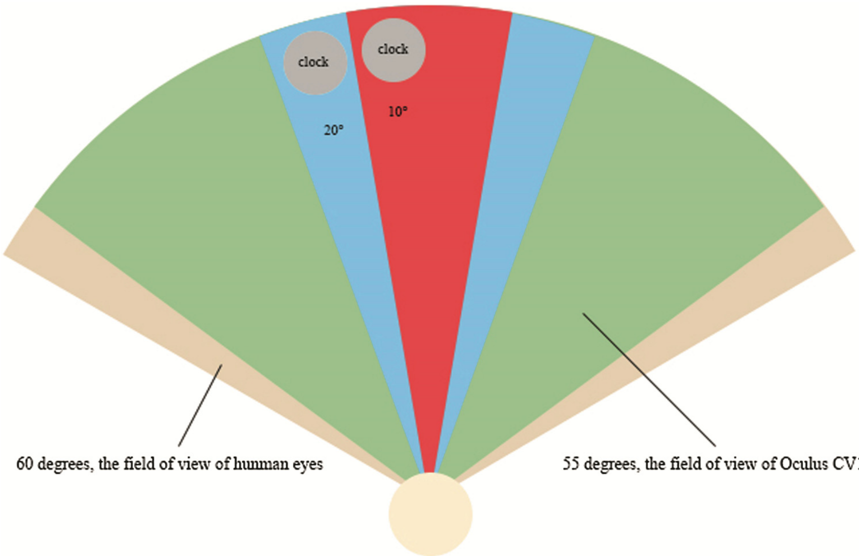


Fig. 7. Simulation of looking at the clock

4 Conclusion

From the experiment results, we found that people's eyesight tends to remain in the central area of the HMD screen when playing a VR game, and the hypothesis was partially supported. As a result, despite the deviations, the direction of the HMD can be used to roughly evaluate a player's focus of attention and to take over the role of an eye tracker to some extent.

Moreover, our work may contribute to the evaluation and optimization of VR games. What has been tested and verified in this paper means that the data directly from HMD rotation sensors can be used to evaluate a user's focus of attention, and maybe further used to evaluation the immersion of the game. This approach employs no equipment other than the HMDs themselves is much more economic and convenient when compared with other methods dependent on expensive devices like in-helmet eye trackers. In the future, the research team also plan to investigate the possibility of using this approach to evaluation the immersion level of VR games based on Charlene Jennett's prior research and findings.

Finally, this the limitations of this experiment are as follow, and future researchers on this topic may need to pay more attention.

- The sample size was not big enough. 20 subjects were not adequate to bring about a universal conclusion.
- Even though people at age 20–26 are the main target users of VR devices, this age range cannot stand for all users of VR games or programs. Future researchers are suggested to expand the age range.

- An in-helmet eye tracker, if affordable, may act as an objective standard reference. Instead of assuming the eyesight focuses at the center of the clock, as we did in this research, it may accurately tell where the focus is, and thus makes the result more accurate.

References

1. Wikipedia: Virtual reality. https://en.wikipedia.org/wiki/Virtual_reality
2. Danaj, M.: Virtual reality. Greater Lansing Bus. Mon. **30**(10), 14–15 (2017)
3. Slater, M.: Presence and the sixth sense. Presence-Teleoperators Virtual Environ. **11**(4), 435–439 (2002). https://doi.org/10.1162/105474602760204327
4. Sanchez-Vives, M.V., Slater, M.: From presence to consciousness through virtual reality. Nat. Rev. Neurosci. **6**, 332–339 (2005). https://doi.org/10.1038/nrn1651
5. Slater, M., Sanchez-Vives, M.V.: Enhancing our lives with immersive virtual reality. Front. Robot. AI **3**, 74 (2016). https://doi.org/10.3389/frobt.2016.00074
6. Chengwei, W.: VR technology and the harmonized HCI environment. J. Comput. Res. Dev. (1997)
7. Hecht, J.: Optical dreams, virtual reality. Opt. Photonics News **27**(6), 24–31 (2016). https://doi.org/10.1364/opn.27.6.000024
8. The official website of STEAM. http://store.steampowered.com/vr/
9. Wikipedia: List of PlayStation VR games. https://en.wikipedia.org/wiki/List_of_PlayStation_VR_games
10. Jennett, C., Cox, A.C., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and defining the experience of immersion in games. Int. J. Hum.-Comput. Stud. **66**(9), 641–661 (2008). https://doi.org/10.1016/j.ijhcs.2008.04.004
11. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990)
12. Agarwal, R., Karahana, E.: Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Q. **24**(4), 665–694 (2000)
13. Duchowski, A.T., Shivashankaraiah, V., Rawls, T., Gramopadhye, A.K., Melloy, B.J., Kanki, B.: Binocular eye tracking in virtual reality for inspection training. Int. J. Ind. Eng.: Theor. Appl. Pract. **9**(2), 89–96 (2000). https://doi.org/10.1145/355017.355031
14. Stengel, M., Grogorick, S.: An affordable solution for binocular eye tracking and calibration in head-mounted displays. In: Proceeding: MM 2015 Proceedings of the 23rd ACM International Conference on Multimedia, pp. 15–24 (2015). https://doi.org/10.1145/2733373.2806265
15. Unity - Manual: Audio listener. https://docs.unity3d.com/Manual/class-AudioListener.html

Get Well Soon! Human Factors' Influence on Cybersickness After Redirected Walking Exposure in Virtual Reality

Julian Hildebrandt[1](✉), Patric Schmitz[2], André Calero Valdez[1], Leif Kobbelt[2], and Martina Ziefle[1]

[1] Human-Computer Interaction Center (HCIC) Chair of Communication Science, RWTH Aachen University, Campus-Boulevard, Aachen, Germany
{hildebrandt,calero-valdez,ziefle}@comm.rwth-aachen.de

[2] Visual Computing Institute, RWTH Aachen University, Mies-van-der-Rohe-Straße, Aachen, Germany
{patric.schmitz,kobbelt}@cs.rwth-aachen.de

Abstract. Cybersickness poses a crucial threat to applications in the domain of Virtual Reality. Yet, its predictors are insufficiently explored when redirection techniques are applied. Those techniques let users explore large virtual spaces by natural walking in a smaller tracked space. This is achieved by unnoticeably manipulating the user's virtual walking trajectory. Unfortunately, this also makes the application more prone to cause Cybersickness. We conducted a user study with a semi-structured interview to get quantitative and qualitative insights into this domain. Results show that Cybersickness arises, but also eases ten minutes after the exposure. Quantitative results indicate that a tolerance towards Cybersickness might be related to self-efficacy constructs and therefore learnable or trainable, while qualitative results indicate that users' endurance of Cybersickness is dependent on symptom factors such as intensity and duration, as well as factors of usage context and motivation. The role of Cybersickness in Virtual Reality environments is discussed in terms of the applicability of redirected walking techniques.

Keywords: Virtual Reality · Cybersickness · Human Factors
Redirected walking · Rotation gain · Immersion

1 Introduction

Since Virtual Reality (VR) systems are becoming a commodity, new challenges occur that focus user experience in addition to technical development. Virtual Environments (VE) are of theoretically infinite size, while VR setups in homely environments are limited in the size of their trackable area. Redirected walking (RDW) techniques are used to overcome this spatial conflict, while maintaining natural walking as the best-perceived navigation metaphor [1]. Using different approaches, they manipulate the user's walking trajectory in the virtual world.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 82–101, 2018.
https://doi.org/10.1007/978-3-319-91581-4_7

Users might, for instance, perceive themselves as walking on a straight line, while in the real world they walk on a curved path. Besides their technical efficiency (i.e. degree of spatial compression), user experience aspects have to be considered when implementing RDW techniques. Symptoms of Cybersickness (e.g. nausea, disorientation, or oculomotor fatigue) pose a crucial threat to VR systems in general and are expected to be even stronger when RDW techniques are applied. Apart from Cybersickness, reduced user immersion can result due to the purposeful degradation of system fidelity [2].

To address the question of how the virtual world can be reliably mapped into real-world boundaries, while keeping immersion as high and Cybersickness as low as possible, we examine the influence of Human Factors on Cybersickness symptoms. We therefore conducted a user study in a VR environment that employs dynamic rotation gains, and measured the influence of various factors on Cybersickness levels before, right after, and 10 min after the VR exposure.

The remainder of this article is structured as follows: We provide an overview on Human Factors in VR, Cybersickness, and RDW techniques in Sect. 2. Our experimental procedure is described in Sect. 3, and the quantitative and qualitative results of the study are presented in Sect. 4. We discuss the role of Cybersickness based on these results and outline future work in Sect. 5. Section 6 summarizes our contribution.

2 Related Work

In this section, we present relevant related work regarding Human Factors in the context of VR applications (Subsect. 2.1), Cybersickness (Subsect. 2.2), and RDW techniques (Subsect. 2.3).

2.1 Human Factors in Virtual Reality

Human Factors are discussed since the early stages of VR development to attain human performance efficiency. Alongside task characteristics (e.g. the suitability of a task for a VE), Stanney et al. identified user characteristics, design constraints imposed by human sensory and motor physiology, as well as health and safety issues that have to be considered when VE systems are implemented [3]. To maximize human performance, the experienced sense of presence should be as high, and the interaction techniques as efficient as possible. Both goals can only be reached if VR hardware provides visual, auditory and haptic/kinesthetic output that convince the sensitive human sensory system of authentic input [4].

Both *presence* and *immersion* are used to describe the relation between the user's (self-)perception in the virtual world and the immediate physical surroundings. There is widespread agreement that both are desirable properties, however their exact definition and distinction is controversially discussed [5]. *Immersion* is either defined as inherent property of the display system that could be objectively quantified [6], or as a perceptual user response to the VE [7]. We adopt the latter definition as the impression of being enveloped by and interacting with a continuous stream of stimuli [7]. We refer to the inherent properties as *system immersion* [2] or *sensory fidelity*, which we define as the degree to which display and transformation of sensory information are similar to the real world [8]. *Presence*, on the other hand, is the subjective experience of being physically situated in the VE instead of the physical locale [7]. We consider system immersion a precondition for (user) immersion, which is in turn a precondition for a sense of presence. Measuring presence is most effectively done by verbal reporting [9] or with a post-questionnaire [7].

2.2 Cybersickness as Critical Factor

Besides human performance and perceptual reactions, the well-being and safety of users is of equal importance [10]. Cybersickness is considered a crucial threat to the adoption of VR technology [3], even if sickness-inducing applications might still be perceived as highly enjoyable [11]. It is mostly defined as symptoms of motion sickness, but occurring during or after VR system exposure [12]. The symptoms vary individually and have multiple influencing factors [13]. They include eye strain, headache, pallor, sweating, dryness of mouth, fullness of stomach, disorientation, vertigo, ataxia, nausea, and might even lead to vomiting [14]. With those symptoms in common, the phenomena motion sickness, simulator sickness and Cybersickness seem equal, but they are caused by slightly different situations: Motion sickness occurs mostly on moving vehicles (e.g. ships) and simulator sickness on devices that simulate motion by visual displays and moving platforms (e.g. flight simulators), while Cybersickness can be induced by visual stimuli alone [15].

A number of different theories attempt to explain the biological mechanisms that cause motion sickness. The *sensory conflict theory* suggests that a mismatch between different sensory subsystems causes the symptoms [16]. It explains factors specific to VR systems well, such as mismatches between user motion and visual display due to latency or tracking inaccuracy. While commonly employed to explain symptoms, the theory has been criticized for it's lack of predictive power [17]. The *postural instability theory* was proposed as an alternative [18]. It explains the symptoms as the body's reaction to phases of unstable posture, during which the postural control loop adapts to unknown or changing environmental conditions. While it also lacks an explanation of the biological causes and effects, in some cases it succeeds in predicting sickness levels based on length and intensity of postural instability phases [19]. Recent research indicates that the theory might be unconfirmed, though no reliable numerical evidence could be obtained [20]. The *rest frame hypothesis* is based on the notion that our perceptual apparatus optimizes the estimation of spatial relationships by assuming certain objects in the environment as stationary [21]. This reference coordinate system is called the rest frame. The theory postulates that sickness symptoms arise due to conflicting cues that indicate different parts of the environment should be considered stationary. Lastly, the *poison theory* attempts to explain why the human body reacts with nausea and vomiting. It argues that perturbations of the spatial frameworks defined by the different senses can be caused by certain types of motion but also the ingestion of toxins [22]. According to the theory, the body reacts with nausea and vomiting as a survival mechanism to remove the alleged poisonous substance.

The most widely applied instrument to quantify Cybersickness is the Simulator Sickness Questionnaire (SSQ) [15], which was originally developed to identify "problematic" flight simulators [23]. The SSQ consists of 16 symptoms that, in a series of factor analyses, were found to cluster in three subscales: nausea, oculomotor, and disorientation. The SSQ was derived from 1,119 underlying data sets of professional pilots using several flight simulators, and scaled such that every subscale and a total score that comprises all 16 symptoms have a zero point and a standard deviation of 15 [13]. Based on this calibration, motion sickness, simulator sickness, and Cybersickness show different symptom profiles: Nausea appears to be highest rated for motion sickness, oculomotor for simulators and disorientation for Cybersickness [24,25]. Furthermore, while Kennedy et al. consider an SSQ total score of 20 as indicating a bad simulator, it turns out that SSQ total scores are higher when applied to VR systems instead of flight simulators [26]. More VR-specific measures are less validated, and therefore not widely adopted, while objective measures (e.g. heart rate, blink rate, electroencephalography) are expensive to perform and rely on intrusive equipment [15].

There are several Human Factors that influence the severity of Cybersickness symptoms. Experience with the used technology is known to have a positive impact on well-being [27], in a way that tolerance towards Cybersickness is learnable, respectively trainable [28]. Furthermore, women tend to be more susceptible to Cybersickness than men. This could be caused by anatomical differences (e.g. women have larger fields of view, which leads to more flicker perception [14]), hormonal levels, or biased response behavior, since men tend to withhold information about vulnerability [29]. Another factor is age: Motion sickness appears stronger on children (2–12 years old), but in contrast to that, Cybersickness appears to be stronger on users older than 30 years. Other factors that contribute to Cybersickness are the overall health status (e.g. overweight, upset stomach, etc.) and mental rotation ability [27]. Further factors depend on task or hardware properties: Low acceleration movements, high degree of control, low time on task, appropriate blur level and low latency have a positive impact on users' well-being [30].

Based on factors that positively influence the occurrence of Cybersickness, a number of technical measures have been proposed to improve the VR system experience. These range from improvements in rendering over specialized sickness-reducing navigation techniques to guidelines for the overall design of virtual scenes. On the rendering level, an adaptive field of view during head rotations [31], dynamic blurring [32], or simulated depth-of-field rendering [30] have been proposed. Navigation methods try to minimize the amount of vection caused by passive movement through the scene [33], or mitigate its impact by providing subtle motion prediction cues [34]. Regarding general scene design it has been shown that scene complexity and realism can increase discomfort in VEs [35] and, e.g., ramps are preferable over stairs [36].

In summary, Human Factors have always been discussed to create VR hardware that provides authentic output (i.e. sensory fidelity) and system immersion. All benchmarks that are used to evaluate VR systems, such as presence, immersion, Cybersickness, and task performance—regardless of their exact definition—benefit from high system immersion, and might therefore be compromised if system immersion is limited.

2.3 Cybersickness in Redirected Walking Applications

Redirected Walking techniques spatially compress the virtual scene, to enable walking through larger virtual spaces than the physical bounds permit. Initial work on the subject demonstrated the feasibility of the method in a restricted scenario [37]. In a virtual scene, subjects walked along a series of predefined waypoints in a zigzag fashion. The turning motion at each waypoint was slowed down, such that by compensating for the reduced rotation, users performed a full 180 degree real-world rotation at each turning point. This way, a large virtual room could be explored while subjects walked back and forth between two ends of the tracked space.

The approach of adaptively increasing or decreasing the effective virtual motion to unnoticeably steer the user onto a desired real-world trajectory is called redirection with *dynamic motion gains*. To generalize the concept for arbitrary VEs, a number of improvements to the method have been proposed. *Steering algorithms* were developed, which compute dynamic motion gains based on universal heuristics [38,39], human motion models [40,41], or path-planning on a set of waypoints [42]. Another, more intrusive approach is the use of *distractors* in the virtual scene. Dynamically moving objects or agents are used to block the user's path and induce a turning motion [43], or to perform scene manipulations while the user is distracted [44,45]. Similarly, the phenomenon of *change blindness* is exploited to perform unnoticed scene manipulations. Examples include changing the placement of doorways when the user is not looking at them to create self-overlapping architecture [46,47] or perform slight scene manipulations during saccades [48] or blinks [49]. Different *perceptual illusions* have also been investigated for unnoticeable scene manipulations [50]. Recent work proposes the use of planar *map folding* that deforms the virtual floor plan to fit into the real-world boundary while maintaining local conformality and bijectivity of the warped space. In the following we focus on RDW techniques that modify the one-to-one mapping of real to virtual motion, either by employing explicit motion gains or by an implicit mapping as with the map folding approach. We consider these particularly prone to Cybersickness due to the sensory conflict between the visual and vestibular organs.

While many results have been achieved regarding the effectiveness of such RDW techniques, the problem of Cybersickness has only been addressed marginally. Most of the cited research does not present results regarding the effect on Cybersickness levels. Where quantitative measures are given, the lack of a control group experiment precludes from attributing the measured levels of Cybersickness to the employed RDW technique instead of the VE itself.

The initial work by Razzaque et al. mentions that subjects in their pilot study ($n = 11$) did not suffer any increase in simulator sickness [37]. The paper itself does not give any numerical evidence for the claim, however. Furthermore, the authors state that increased Cybersickness should not be caused, because "the technique keeps the visual, auditory and vestibular cues consistent". We object to that notion, in so far that the application of motion gains *does* cause an inconsistency between visual and vestibular cues, which we believe to be a major cause of Cybersickness symptoms. Later work restates the claim that additional Cybersickness caused by RDW techniques should not be significant for VR applications [51]. The sample size is derived for a power analysis to support that claim, which for the best case (an assumed SSQ population mean of 11 and an effect size of 5) requires $N = 266$ samples. Since at the time there was no capacity to conduct a study of that size, the claim remains unproven.

The notable work by Steinicke et al. on detection thresholds of motion gains does provide some measured SSQ data [52]. In the user study ($n = 14$), increased Cybersickness levels are reported, however from the given data it is not decidable whether the change was significant. The authors state that a follow-up experiment was conducted among subjects with "high Post-SSQ scores", but do not give any details about the experimental procedure. Another consideration with the reported values is that an SSQ was filled out once before and once after a series of three experiments. Although not explicitly stated, the order of experiments in the paper suggests that the experiment on rotation gain (E1) was performed first, and the Post-SSQ score was taken after all three experiments were completed, which took three hours overall according to the authors. Since rotation gains in particular have a strong effect on Cybersickness [33,53], it is possible that intermediate SSQ scores would have been much higher than the final Post-SSQ score suggests.

3 Method

In the following, the experimental design, the variables, the sample, as well as the analysis procedure are described.

3.1 Experiment Design

The experiment design was equivalent to the design described in Schmitz et al. [54]. Participants completed a preliminary questionnaire and completed four trials in a testbed environment while being redirected by dynamic rotational gains. A trial consisted of a target-collection task that had to be performed until the individual *threshold of limited immersion* [54] was reached. Subjects completed two trials in increasing and decreasing condition each, if they did not abort the experiment. Every participant was informed beforehand that undesirable side effects might occur and that they should not continue the experiment if they get uncomfortable, but they were uninformed about the applied redirection technique. Participants with a medical history of epilepsy were excluded for safety reasons. Figure 1 shows an exemplary view on our testbed environment.

Fig. 1. View of the subject; 2 pillars are out of sight.

After the experiment, participants completed a post-questionnaire and a semi-structured guideline-based interview. The Simulator Sickness Questionnaire (SSQ) was completed at three times during the procedure: immediately before participants entered the VE (SSQpre), immediately after leaving the VE (SSQ-post) and approximately 10 min later (SSQfinal).

3.2 Variables

We surveyed the following Human Factors in the preliminary questionnaire: gender, age, education degree, a VZ-2 paperfolding test as measure of mental rotation ability [55], self-efficacy towards technology (SET) [56], and the tendency to be immersed [7]. Furthermore we developed Likert scales to measure the subjective overall health condition, the experience with VR technology, and the tolerance for nauseous activities (e.g. riding a roller-coaster). In addition, we asked for self-reports on the own sense of direction and on overweight using a single Likert item each.

As evaluation criteria we surveyed perceived presence [7] and perceived immersion [57] as VR-related dimensions; enjoyment and anxiousness [58] as perceived emotions, as well as the behavioral intention scale of UTAUT2 as measure of technology adoption [59]. Furthermore, we operationalized trust in redirected walking (e.g. *"I feared to touch real objects or walls"* [negative item]) and the perception of the motion tracking as convincing (e.g. *"It felt strange to move around in the virtual environment"* [negative item]) in 5 items each.

3.3 Analysis Procedure

Self-reporting measures were transformed into pseudo interval scales if Cronbach's-alpha was above 0.5. SSQ total scores and sub-scores were calculated by summing up the symptom ratings and multiplying them with the according factors [23]. Since there is no interpretative meaning of SSQ score distances defined, we rely on non-parametric methods for statistical analysis.

Qualitative material was recorded and transcribed according to the well-established GAT2 system [60]. Those transcriptions were our sampling units. We define the answer on the question *"Under which circumstances would you endure general discomfort after VR usage?"*, as well as all further queries by the examiner to fully understand a answer recording unit. We defined a single phrase as smallest possible content unit and the whole answer as the context unit. To achieve high inter-coder agreement, the procedural approach of consensual coding was conducted [61]. We defined four main categories as initial category system: *Unmitigated preferences*, *application factors*, *symptom factors* and *other*. Four professionals took part in the coding procedure and assigned content units into those main categories independently as a first step. As a second step, they defined sub-categories that were determined inductively. Based on these four category systems, a consensual category system was elaborated and (re-)defined by all four coders. As a conclusive step, the overall material was assigned into this consensual system. The final results are unanimous.

3.4 Sample Description

Overall, 52 participants (50% female) took part in our study. The age of the participants ranged between 19 and 35 years ($M = 24.33$; $SD = 3.18$). 48% of the sample stated to own a high-school diploma, the other 52% reported to have obtained a higher education degree. In total, 12 participants (23.1%) aborted the experiment before reaching the final condition because of Cybersickness.

4 Results

In this section we describe quantitative and qualitative results of our study. We start by describing the dependent variables—SSQ nausea, oculomotor, discomfort and total score—as well as their inter-correlations. We further analyze Human Factors' influences on subjective evaluation criteria. The last subsection describes the results of the qualitative text analysis.

4.1 Quantitative Effects on Cybersickness

Description of SSQ Scales. Table 1 shows descriptive statistics of the SSQ values for each time of measurement, as well as their differences in total scores and inter-correlations between the sub-scales nausea, oculomotor and disorientation at corresponding times. Since all sub-scales are heavily intercorrelated, we narrow down the further analysis to the interpretation of total scores. The average participant came to the experiment with $M = 16.18$ ($SD = 15.99$) points on SSQ total score, experienced a decrease in well-being to $M = 34.22$ ($SD = 39.51$)

points and recovered to $M = 19.44$ ($SD = 20.91$) points 10 min after the experiment. Right after the experiment, there was no participant without any symptom of cybersickness at all (Min = 3.74, $M = 50.05$, $SD = 44.80$), and participants left with $M = 30.29$ ($SD = 39.58$) points on average. Furthermore, all total scores are intercorrelated: Higher symptoms before the VE exposure are significantly related to higher SSQ scores right after ($\tau = .41**, p < .01$) and higher SSQ scores 10 min after the experiment ($\tau = .40**, p < .01$). In addition, the two SSQ scores after the exposure are also positively associated ($\tau = .74**, p < .01$).

Table 1. Descriptives and Kendall-τ Intercorrelations for SSQ Subscales: TS = Total Score, N = Nausea, O = Oculomotor, D = Disorientation.

		Min	Max	M	SD	resp. Nausea	resp. Oculomotor	resp. Disorientation
SSQpre	Nausea	0	50.88	10.72	12.36	-		
	Oculomotor	0	90.96	18.42	19.67	.42**	-	
	Disorientation	0	64.96	11.4	14.89	.32**	.47**	-
	Total score	0	71.53	16.18	15.99	.59**	.82**	.60**
SSQpost	Nausea	0	184.44	42.31	38.83	-		
	Oculomotor	0	146.55	36.6	33.46	.63**	-	
	Disorientation	0	296.96	57.91	60.46	.68**	.65**	-
	Total score	3.74	222.00	50.05	44.80	.78**	.80**	.80**
SSQfinal	Nausea	0	146.28	25.18	33.45	-		
	Oculomotor	0	136.44	23.44	30.03	.68**	-	
	Disorientation	0	185.60	34.18	49.32	.75**	.66**	-
	Total score	0	172.67	30.29	39.58	.84**	.81**	.82**
$\Delta SSQpost - SSQpre$	**Total score**	-14.8	207.20	34.22	39.51	n/a	n/a	n/a
$\Delta SSQfinal - SSQpost$	**Total score**	-86.3	29.60	-19.44	20.91	n/a	n/a	n/a

Note: Correlation columns show the sub-scales that correspond to the measuring time of the row.
$^{*}p < .05$, $^{**}p < .01$.

The change in Cybersickness scores is quite visible in every sub-scale of the SSQ (see Fig. 2). The change in SSQ total score is significant according to Friedman's Anova ($\chi^2(2) = 51.61$, $p < .01$). Wilkoxon tests in pairwise comparison mode were used to follow up this finding. The SSQ score changed significantly from pre to post ($T = -1.25, p < .01$) and from post to final ($T = 1.12, p < .01$), but pre and final were not significantly different ($T = -.13, p = .51, n.s.$). A power-analysis was conducted on this last finding, yielding $1 - \beta = 0.55$. We therefore are not able to generalize difference or equality based on our sample and effect size.

Effects of Human Factors. When considering Human Factors, there are several significant correlations with SSQ scores (see Table 2). In line with theory, female gender is positively associated to Cybersickness, which is indicated by positive correlations between gender (dummy coded) and SSQpost ($\tau = .26*, p < .05$), SSQfinal ($\tau = .26*, p < .05$) and ΔPrePost ($\tau = .23*, p < .05$). A Mann-Whitney-U-test confirms this gender effect: Men were more resilient to symptoms of Cybersickness than women right after the experiment ($U = 461.50, p < .05$)

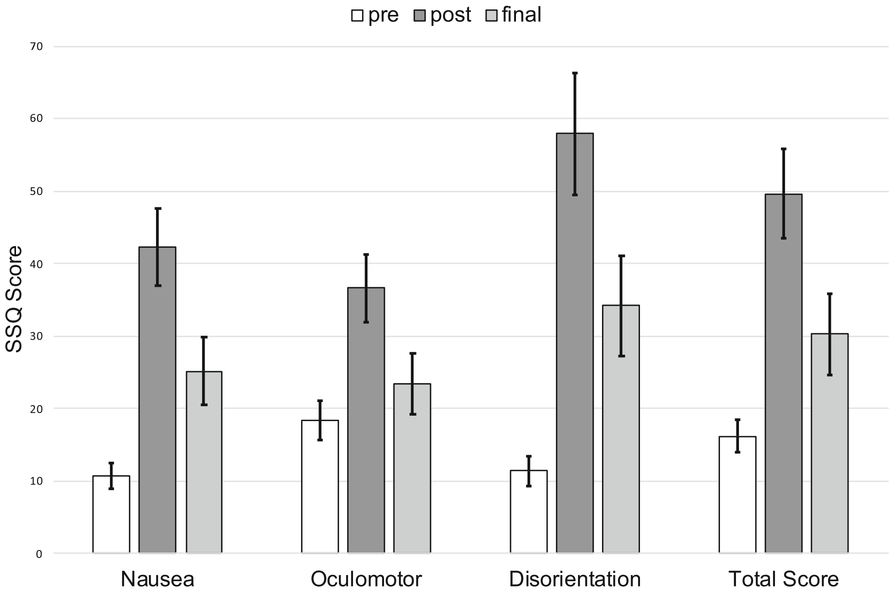


Fig. 2. Sub-scales and total score of SSQ among all times of measurement $\pm SE$. Important Note: The y-axis is cropped for readability reasons and every scale has a different theoretical maximum.

and 10 min after the experiment ($U = 425.50, p < .05$). Furthermore, the increase in Cybersickness between men and women caused by the experiment was higher for women than for men ($U = 430.50, p < .05$). Other Human Factors, that are associated to gender, are also correlated to SSQpost and SSQfinal: With lower SET comes higher SSQpost ($\tau = -.21*, p < .05$) and higher SSQfinal ($\tau = -.26*, p < .05$), the same applies to a low perceived sense of direction ($\tau = -.23*, p < .05$, resp. $\tau = -.24*$, $p < .05$). Other significant correlations with SSQ scores can be found between age and symptoms of Cybersickness before the experiment ($\tau = .23*, p < .05$), with younger participants having less symptoms. Apart from that, a better overall health condition was negatively associated to the SSQ score 10 min after the VR exposure ($\tau = -.26*, p < .05$).

Effects on Evaluation. Regarding the subjective evaluation of the VR experience, several of the surveyed dimensions were associated to Cybersickness (see Table 3). Subjects who suffered more from Cybersickness right after leaving the VE experienced less presence ($\tau = -.19*, p < .05$) and less immersion ($\tau = -.21*, p < .05$), but these associations were not significant at the final measure. Additionally, a lower perception of enjoyment ($\tau = -.23*, p < .05$, resp. $\tau = -.28**, p < .01$) and a higher valuation of the motion as convincing ($\tau = -.25*, p < .05$, $\tau = -.28**, p < .05$) was associated to higher SSQ scores at both measuring times after the VR task. Surprisingly, a higher sense of anxiousness ($\tau = .43**, p < .01$, resp. $\tau = .45**, p < .01$) and a lower

Table 2. Kendall-τ correlations between SSQ total score and human factors. $p* < .05$, $p** < .01$.

	α	M	SD	SSQpre	SSQpost	SSQfinal	ΔPrePost	ΔPostFinal
Gender	n/a	n/a	n/a		.26*	.26*	.23*	
Age	n/a	24.3	3.18	.23*				
Education	n/a	n/a	n/a					
VZ-2 paperfolding	n/a	14.5	3.74					
HealthCondition	0.59	4.82	0.78			−.26*		
Experience with VR	0.78	2.95	0.66					
Tolerance for nauseous activities	0.74	3.96	1.16					
SET	0.83	4.26	0.83		−.21*	−.26*		
Immersion tendency	0.62	4.08	0.73					
Perceived sense of direction	n/a	3.60	1.34		−.23*	−.24*		
Distance to standard weight	n/a	0.82	0.96					

Note: $*p < .05$, $**p < .01$, *Gender was dummy-coded with 1 = male, 2 = female for correlation.*

perception of trust ($\tau = -.27**, p < .01$, resp. $\tau = -.25**, p < .01$) was not just correlated to higher SSQ scores at both measuring times after, but also to the SSQ score even before entering the VE ($\tau = -.23*, p < .05$).

To further examine these findings, we calculated partial correlation: When we control SSQpre on the relationship between anxiousness and SSQpost resp. SSQfinal, we find the partial correlations to be still significant ($r = .51**, p < .01$, resp. $r = .40**, p < .01$). This does not apply to the relationship between trust and these SSQ scores ($r = -.26, p = .06, n.s.$, resp. $r = -.19, p = .18, n.s.$).

Moreover, we found significant correlations between the first change in Cybersickness (ΔPrePost) and all surveyed dimensions except immersion and technology adoption: A higher increase in Cybersickess is associated to lower presence ($\tau = -.23*, p < .05$), lower enjoyment ($\tau = -.26*, p < .05$), lower trust ($\tau = -.24*, p < .05$), lower valuation of the motion as convincing ($\tau = -.21*, p < .05$) and higher anxiousness ($\tau = .40**, p < .01$). Furthermore, technology adoption was negatively associated to the final SSQ score ($\tau = -.20*, p < .05$).

Table 3. Kendall-τ correlations between SSQ total score and evaluation criteria.

	α	M	SD	SSQpre	SSQpost	SSQfinal	ΔPrePost	ΔPostFinal
Presence	0.69	2.86	0.53		−.19*		−.23*	
Immersion	0.66	3.49	0.88		−.21*			
Enjoy	0.75	2.18	0.99		−.23*	−.28**	−.26**	
Anxiousness	0.74	1.46	0.96	.23*	.43**	.45**	.40**	−.22*
Technology adoption	0.86	2.86	1.54			−.20*		
Trust	0.69	3.12	1.01	−.21*	−.27**	−.25*	−.24*	
Convincing movement	0.58	2.34	1.00		−.25*	−.28**	−.21*	

Note: $*p < .05$, $**p < .01$.

4.2 Qualitative Role of Cybersickness

To complement our quantitative results, we report qualitative results in this section. Overall, we were able to identify 105 content units that could be assigned in the following categories.

Main Category 1: Unmitigated Preferences. The sampling unit had to include an expression that Cybersickness would (or would not) be endured unexceptionally, to be assigned into this main category. We inductively developed the obvious subcategories *Unmitigated Endurance* and *Unmitigated Refusal.*

Unmitigated Endurance. Only one content unit was covered by this category. The subject stated that she would "definitively use VR" even if symptoms occur, because she always endured the symptoms in previous VR experiences.

Unmitigated Refusal. These units contained several arguments that were used to justify the answer, but we could not find any pattern that would justify the creation of another category layer. Subjects explained their refusal with the basic need to feel fine or stated that they would immediately abort the VR experience if Cybersickness arises. Others argued that they would not be capable to focus on the application anymore and some participants exclaimed that Cybersickness would just be "a no-go!" without any further justification.

Main Category 2: Effect of Application Context. Being way more relevant to the research question than the first category, units had to contain statements that put endurance of Cybersickness in relation to the context of the VR-Application to be assigned in this category. We inductively developed three subcategories: *Exciting Application*, *Serious Application* and *Extrinsic Motivation.*

Exciting Application. Subjects argued that they would endure Cybersickness if the "fun-factor predominates the symptoms", or that they would endure it just like they endure motion sickness after a roller-coaster. The VR experience should be "exciting", "extremely innovative" or just an "intense video game" to compensate for the symptoms. Interestingly, some subjects answered the question in a way that they believe they would not even recognize Cybersickness if the application is entertaining enough.

Serious Application. In contrast to the previous subcategory, some content units addressed serious applications instead of fun. Subjects argued that the VR experience should offer "some serious benefit" or "serve a good purpose". Three units mentioned very specific applications: Subjects would endure Cybersickness, if they would need to practice surgical operations in VR or if they could "quench their thirst for knowledge" by visiting the Louvre in Paris. Another unit contains the argument that VR could be necessary for work and adverse effects should therefore be endured.

Extrinsic Motivation. Two participants mentioned extrinsic motivation as a condition to endure Cybersickness without mentioning specific application domains. They argued that they would tolerate discomfort if "they have to use the application" or for reasons of conscientiousness.

Main Category 3: Effect of Cybersickness Characteristics. This main category contains units that include characteristics of the Cybersickness symptoms itself. We elaborated the category system by determining two subcategories: *Duration* and *Intensity*.

Duration. Content units in this subcategory included arguments regarding the moment of occurrence or the duration of the symptoms. Subjects argued that they would endure Cybersickness if the symptoms were only present as long as the VE exposure lasts and not longer. Other subjects stated that the symptoms should not last longer than two hours, or that the symptoms should not occur continually, but only in singular peaks.

Intensity. Subjects also stated that the intensity of the symptom should remain under a certain threshold. Cybersickness should be "still bearable" and "not too extreme". Another subject expressed the wish to examine the point, where she "can't take it anymore" and was optimistic "that it probably will get better after a certain habituation phase".

Main Category 4: Other Effects. We found some content units that were not assignable in the previous categories and assigned them in three sub-categories: *Information about Cybersickness*, *Habituation* and *Cybersickness as Part of User Experience*.

Information About Cybersickness. Two subjects stated that they would endure Cybersickness if they were well-informed that Cybersickness occurs at all, and furthermore can be sure that the reason for the symptoms "is just the algorithm and nothing serious".

Habituation. Some subjects stated that they would endure the symptoms if they get the opportunity to experience Cybersickness more often to get used to the symptoms and to experience their individual threshold of discomfort.

Cybersickness as Part of User Experience. Other units contained some interesting meta-aspects: Cybersickness might be accepted if the purpose of the VR application is to make the user resilient against Cybersickness, or even that Cybersickness might be part of the VE itself: Applications could deliberately induce Cybersickness to simulate drunk-driving and therefore raise awareness for responsible behavior in traffic.

Additional Findings. From a user experience perspective, it is noteworthy that subjects often used phrases like "I can't think of a specific application right now" or "I can't think of anything other than video games as applications". Apparently, users—even though younger and technology prone users volunteered in the study— still have difficulties to envision usage scenarios outside of their experience with VR applications and to contribute to user-centered requirements analysis of future applications.

5 Discussion

We conducted a user study where participants experienced a virtual environment with rotation gains as RDW technique, and measured Cybersickness with the well-established SSQ before, right after, and 10 min after the exposure. Results show that Cybersickness indeed arises, but also eases significantly after 10 min. Due to the size of the experiment, we can not conclude whether the symptoms reach the same level as before—the power of the experiment was only $1-\beta = .55$. From other studies using VR and RDW we know that Cybersickness decreases quickly for some of the participants, yet others report of persisting symptoms for several hours. Apparently, there are considerable individual differences in how persons react to VR applications with Cybersickness symptoms. In this context, further research is required on Human Factors' influence on the recovery rate of Cybersickness as well as an understanding of the different responsiveness of users to sickness symptoms.

In contrast to the huge body of knowledge that individual variables and user diversity are central factors influencing the behaviors of and the attitudes towards novel technology [62–65], user factors showed only a marginal influence in this VR experiment. On the basis of the current data we can not conclusively explain the reasons for the small influence of individual variance. Speculating, two major arguments can be referred to in this context. On the one hand we only had a small and homogeneous sample, as all participants were young, highly educated, healthy, and technology savvy. Thus, the effect of user diversity could have been veiled as the differences were not detectable. On the other hand, one could speculate that the experience of rotational gains in VR environments might be unique for all participants in terms of physical reactions, and behaviors. Future studies will have to replicate the findings with more, and more diverse, participants.

However, we could replicate previous findings that women tend to be more susceptible to symptoms of Cybersickness than men at both times of measurement after the VR exposure. Obviously, women seem to be more responsive to physical and/or perceptual distortion effects. However, the higher sensitivity of women could also be a reporting bias. It is known from previous research that women are more sensible towards perception of bodily experiences, in line with a lower tolerance of pain sensations [66–68]. In addition to gender differences in physical perceptions, women could also be more open-minded to communicate somato-sensations and physical (dis-)comfort in contrast to men. On the base of the current data, we can not decide which of the two explanations, the physical (higher sensitivity to bodily stimuli) or the socio-cognitive explanation (a lower threshold to communicate bodily stimuli) might account for the gender differences in perception and/or communication of Cybersickness symptoms. While this distinction might be insightful from a psycho-physiological perspective, it is not relevant from a Human Factors perspective: Whenever there are gender differences in perception and acceptance of technology, technical designers need to consider those differences due to the claim of technology designs for all and universal access.

Furthermore, Cybersickness appeared to be stronger on subjects with low self-efficacy towards technology and subjects with low perceived sense of direction—which however are predominately women. Complementing the qualitative indication that Cybersickness was less relevant if the symptoms could be attributed directly to the redirection technique, we conclude that information about the specific applied redirection technique could be beneficial. Users who tend to feel insecure about technology might profit considerably from knowing that their symptoms are no reason to worry, but a result of their higher sensitiveness to sensory stimuli.

SSQ scores are highly subjective and even if it would be possible to quantify e.g. nausea objectively, the users' evaluation would still address the subjective perception. This points out a fundamental challenge for research on motion sickness, simulator sickness, and Cybersickness that uses the SSQ as a measure. While it is highly relevant to define a score that reflects the quality of a flight simulator for pilots, it is also highly domain specific with respect to both user and technology. The SSQ measure is therefore not necessarily suitable to evaluate the effects of Cybersickness in general VR applications for untrained users. A more reliable and generally applicable operationalization of Cybersickness is needed, possibly incorporating additional factors, such as objective physiological measures and symptoms that are not derived from motion sickness (e.g. impression of standing next to oneself or foreign body feeling caused by virtual body parts—out-of-body experiences).

In line with work by von Mammen et al. [11], our qualitative data raises the question of the relevance of Cybersickness as a predictor of the acceptance of VR. Similar to experiences such as roller-coaster rides, users conduct a trade-off evaluation of the benefits (e.g. fun) versus the potential costs (e.g. discomfort). This trade-off applies to professional contexts as well, as workers tend to go to work even if they feel ill. The endurance of Cybersickness could therefore also depend on the context and motivation to use VR. It would be highly relevant to understand how different users solve these trade-offs in distinct contexts. Furthermore, since we know that Cybersickness symptoms decrease with VR experience, future work should address to what extend this robustness is trainable. Human Factors that influence training progress should be investigated. The development of accepted VR applications that serve as Cybersickness habituation environments needs to be explored. Based on our data, such applications should not only be exiting, but also highly informative about possible VR benefits in different contexts as well as RDW techniques and the causes of Cybersickness in general.

6 Conclusion

We conducted a user study on the influence of Human Factors on Cybersickness levels in a VR environment that uses motion gains as a redirection technique. Participants were instructed to collect virtual pillars as long as they perceive their motion as natural, while we increased resp. decreased the rotation gain in discrete steps on every target collection. The experiment was finished after two conditions with decreasing gain and two conditions with increasing gain in randomized order. Before, right after, and 10 min after the VR-exposure, we measured symptoms of Cybersickness using the well-established SSQ. Furthermore we quantified demographic data, self-efficacy towards technology, ability of mental rotation, overall health condition, subjective sense of orientation, experience with 3D technology, tendency to be immersed, and tolerance for nausea inducing activities. Moreover, we measured perceived immersion, presence, trust, technology adoption, perception of realistic movement, enjoyment, and anxiousness as criteria of evaluation. The role of Cybersickness was furthermore examined via post-experimental qualitative interviews.

We came to the conclusion that Cybersickness and limited immersion are thresholds to be considered when implementing algorithms for redirected walking. Furthermore, future studies should investigate the influence of usage contexts on the acceptance of Cybersickness, as well as experimental paradigms to measure the intensity of rotational manipulation on which Cybersickness exceeds the acceptable threshold. In a final analysis, the cross-over effects when more than one redirection technique is used should be examined in order to implement an integrated approach that delivers high spatial compression and rich user experience.

Acknowledgments. We would like to thank Susanne Gohr, Fabian Comans and Jennifer Kirstgen for sharing their expertise in content analysis by participating in consensual coding. Furthermore, we would like to thank the strategy funds of RWTH Aachen University within the project house "ICT Foundations of a Digitized Economy, and Society" for funding this work. Special thanks go to Ming Li as well as Torsten Kuhlen for their huge inspiration in early stages of the project.

References

1. Usoh, M., Arthur, K., Whitton, M.C., Bastos, R., Steed, A., Slater, M., Brooks Jr., F.P.: Walking > walking-in-place > flying, in virtual environments. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 359–364. ACM Press/Addison-Wesley Publishing Co. (1999)
2. Nilsson, N.C., Nordahl, R., Serafin, S.: Immersion revisited: a review of existing definitions of immersion and their relation to different theories of presence. Hum. Technol. **12**, 108–134 (2016)
3. Stanney, K.: Realizing the full potential of virtual reality: human factors issues that could stand in the way. In: Proceedings of Virtual Reality Annual International Symposium, pp. 28–34. IEEE (1995)
4. Wann, J., Mon-Williams, M.: What does virtual reality need?: human factors issues in the design of three-dimensional computer environments. Int. J. Hum. Comput. Stud. **44**(6), 829–847 (1996)
5. Baños, R.M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., Rey, B.: Immersion and emotion: their impact on the sense of presence. CyberPsychol. Behav. **7**(6), 734–741 (2004)
6. Slater, M.: A note on presence terminology. Presence Connect **3**(3), 1–5 (2003)
7. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence **7**(3), 225–240 (1998)
8. Bystrom, K.E., Barfield, W., Hendrix, C.: A conceptual model of the sense of presence in virtual environments. Presence: Teleoper. Virtual Environ. **8**(2), 241–244 (1999)
9. Slater, M., Steed, A.: A virtual presence counter. Presence: Teleoper. Virtual Environ. **9**(5), 413–434 (2000)
10. Stanney, K.M., Mourant, R.R., Kennedy, R.S.: Human factors issues in virtual environments: a review of the literature. Presence **7**(4), 327–351 (1998)
11. von Mammen, S., Knote, A., Edenhofer, S.: Cyber sick but still having fun. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, pp. 325–326. ACM (2016)
12. Mousavi, M., Jen, Y.H., Musa, S.N.B.: A review on cybersickness and usability in virtual environments. In: Advanced Engineering Forum, vol. 10, pp. 34–39. Trans Tech Publications Ltd. (2013)
13. Kennedy, R.S., Fowlkes, J.E.: Simulator sickness is polygenic and polysymptomatic: implications for research. Int. J. Aviat. Psychol. **2**(1), 23–38 (1992)
14. LaViola Jr., J.J.: A discussion of cybersickness in virtual environments. SIGCHI Bull. **32**(1), 47–56 (2000)
15. Davis, S., Nesbitt, K., Nalivaiko, E.: A systematic review of cybersickness. In: Proceedings of the 2014 Conference on Interactive Entertainment, pp. 1–9. ACM (2014)
16. Reason, J.T., Brand, J.J.: Motion Sickness. Academic Press, Cambridge (1975)
17. McCauley, M.E., Sharkey, T.J.: Cybersickness: perception of self-motion in virtual environments. Presence: Teleoper. Virtual Environ. **1**(3), 311–318 (1992)
18. Riccio, G.E., Stoffregen, T.A.: An ecological theory of motion sickness and postural instability. Ecol. Psychol. **3**(3), 195–240 (1991)
19. Stoffregen, T.A., Smart Jr., L.J.: Postural instability precedes motion sickness. Brain Res. Bull. **47**(5), 437–448 (1998)
20. Dennison, M.S., D'Zmura, M.: Cybersickness without the wobble: experimental results speak against postural instability theory. Appl. Ergon. **58**, 215–223 (2017)

21. Prothero, J.D.: The role of rest frames in vection, presence and motion sickness (1998)
22. Treisman, M.: Motion sickness: an evolutionary hypothesis. Science **197**(4302), 493–495 (1977)
23. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993)
24. Rebenitsch, L., Owen, C.: Review on cybersickness in applications and visual displays. Virtual Reality **20**(2), 101–125 (2016)
25. Kennedy, R.S., Drexler, J.M., Compton, D.E., Stanney, K.M., Lanham, D.S., Harm, D.L.: Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences. In: Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues, p. 247 (2003)
26. Stanney, K.M., Kennedy, R.S., Drexler, J.M.: Cybersickness is not simulator sickness. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 41, pp. 1138–1142. SAGE Publications, Los Angeles (1997)
27. Kolasinski, E.M.: Simulator sickness in virtual environments. Technical report, Army Research Institute for the Behavioral and Social Sciences, Alexandria (1995)
28. Stanney, K.M., Hale, K.S., Nahmens, I., Kennedy, R.S.: What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. Hum. Factors **45**(3), 504–520 (2003)
29. Biocca, F.: Will simulation sickness slow down the diffusion of virtual environment technology? Presence: Teleoper. Virtual Environ. **1**(3), 334–343 (1992)
30. Porcino, T.M., Clua, E., Trevisan, D., Vasconcelos, C.N., Valente, L.: Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–6. IEEE (2017)
31. Fernandes, A.S., Feiner, S.K.: Combating VR sickness through subtle dynamic field-of-view modification. In: 2016 IEEE Symposium on 3D User Interfaces (3DUI), pp. 201–210. IEEE (2016)
32. Budhiraja, P., Miller, M.R., Modi, A.K., Forsyth, D.: Rotation blurring: use of artificial blurring to reduce cybersickness in virtual reality first person shooters. arXiv preprint arXiv:1710.02599 (2017)
33. Kemeny, A., George, P., Mérienne, F., Colombet, F.: New VR navigation techniques to reduce cybersickness. Electron. Imaging **2017**(3), 48–53 (2017)
34. Jeng-Weei Lin, J., Parker, D., Lahav, M., Furness, T.: Unobtrusive vehicle motion prediction cues reduced simulator sickness during passive travel in a driving simulator. Ergonomics **48**(6), 608–624 (2005)
35. Davis, S., Nesbitt, K., Nalivaiko, E.: Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. In: Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015), vol. 27, p. 30 (2015)
36. Dorado, J.L., Figueroa, P.A.: Ramps are better than stairs to reduce cybersickness in applications based on a HMD and a gamepad. In: 2014 IEEE Symposium on 3D User Interfaces (3DUI), pp. 47–50. IEEE (2014)
37. Razzaque, S., Kohn, Z., Whitton, M.C.: Redirected walking. In: Proceedings of EUROGRAPHICS, vol. 9, pp. 105–106. Eurographics Association, Manchester (2001)
38. Field, T., Vamplew, P.: Generalised algorithms for redirected walking in virtual environments. In: AISAT 2004 International Conference on Artificial Intelligence in Science and Technology, pp. 58–63. University of Tasmania (2004)

39. Hodgson, E., Bachmann, E.: Comparing four approaches to generalized redirected walking: simulation and live user data. IEEE Trans. Vis. & Comput. Graph. (TVCG) **19**(4), 634–643 (2013)
40. Hutton, C., Suma, E.: A realistic walking model for enhancing redirection in virtual reality. In: Virtual Reality Conference (VR), pp. 183–184. IEEE (2016)
41. Zhang, R., Kuhl, S.A.: Flexible and general redirected walking for head-mounted displays. In: VR, pp. 127–128. IEEE (2013)
42. Azmandian, M., Yahata, R., Bolas, M., Suma, E.: An enhanced steering algorithm for redirected walking in virtual environments. In: Virtual Reality Conference (VR), pp. 65–66. IEEE (2014)
43. Peck, T.C., Fuchs, H., Whitton, M.C.: Improved redirection with distractors: a large-scale-real-walking locomotion interface and its effect on navigation in virtual environments. In: Virtual Reality Conference (VR), pp. 35–38. IEEE (2010)
44. Peck, T.C., Fuchs, H., Whitton, M.C.: Evaluation of reorientation techniques and distractors for walking in large virtual environments. IEEE Trans. Vis. Comput. Graph. **15**(3), 383–394 (2009)
45. Chen, H., Fuchs, H.: Supporting free walking in a large virtual environment: imperceptible redirected walking with an immersive distractor. In: Proceedings of the Computer Graphics International Conference, p. 22. ACM (2017)
46. Suma, E.A., Lipps, Z., Finkelstein, S., Krum, D.M., Bolas, M.: Impossible spaces: maximizing natural walking in virtual environments with self-overlapping architecture. IEEE Trans. Vis. Comput. Graph. **18**(4), 555–564 (2012)
47. Vasylevska, K., Kaufmann, H., Bolas, M., Suma, E.A.: Flexible spaces: dynamic layout generation for infinite walking in virtual environments. In: 2013 IEEE Symposium on 3D User Interfaces (3DUI), pp. 39–42. IEEE (2013)
48. Bolte, B., Lappe, M.: Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. IEEE Trans. Vis. Comput. Grap. (TVCG) **21**(4), 545–552 (2015)
49. Langbehn, E., Bruder, G., Steinicke, F.: Subliminal reorientation and repositioning in virtual reality during eye blinks. In: Proceedings of the 2016 Symposium on Spatial User Interaction, p. 213. ACM (2016)
50. Steinicke, F., Bruder, G.: Using perceptual illusions for redirected walking. IEEE Comput. Graph. Appl. **33**(1), 6–11 (2013)
51. Razzaque, S.: Redirected walking. University of North Carolina at Chapel Hill (2005)
52. Steinicke, F., Bruder, G., Jerald, J., Frenz, H., Lappe, M.: Estimation of detection thresholds for redirected walking techniques. IEEE Trans. Vis. Comput. Graph. **16**(1), 17–27 (2010)
53. Trutoiu, L.C., Mohler, B.J., Schulte-Pelkum, J., Bülthoff, H.H.: Circular, linear, and curvilinear vection in a large-screen virtual environment with floor projection. Comput. Graph. **33**(1), 47–58 (2009)
54. Schmitz, P., Hildebrandt, J., Valdez, A.C., Kobbelt, L., Ziefle, M.: You spin my head right round: threshold of limited immersion for rotation gains in redirected walking. IEEE Trans. Vis. Comput. Graph. **24**(4), 1623–1632 (2018)
55. Ekstrom, R.B., Dermen, D., Harman, H.H.: Manual for Kit of Factor-Referenced Cognitive Tests, vol. 102. Educational Testing Service Princeton, Princeton (1976)
56. Beier, G.: Kontrollüberzeugungen im Umgang mit Technik: ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme. dissertation. de (2003)

57. Tcha-Tokey, K., Christmann, O., Loup-Escande, E., Richir, S.: Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. Int. J. Virtual Reality **16**(1), 33–48 (2016)
58. Pekrun, R., Goetz, T., Frenzel, A.C., Barchfeld, P., Perry, R.P.: Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). Contemp. Educ. Psychol. **36**(1), 36–48 (2011)
59. Venkatesh, V., Thong, J.Y., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**, 157–178 (2012)
60. Selting, M., Auer, P., et al.: A system for transcribing talk-in-interaction: gat 2 translated and adapted for english by elizabeth couper-kuhlen and dagmar barth-weingarten. Gesprächsforschung-Online-Zeitschrift zur verbalen Interaktion **12**, 1–51 (2011)
61. Kuckartz, U.: Qualitative Text Analysis: A Guide to Methods, Practice and Using Software. Sage, Thousand Oaks (2014)
62. Arning, K., Ziefle, M.: Understanding age differences in PDA acceptance and performance. Comput. Hum. Behav. **23**(6), 2904–2927 (2007)
63. Li, M., Arning, K., Vervier, L., Ziefle, M., Kobbelt, L.: Influence of temporal delay and display update rate in an augmented reality application scenario. In: Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia, pp. 278–286. ACM (2015)
64. Brell, J., Calero Valdez, A., Schaar, A.K., Ziefle, M.: Gender differences in usage motivation for social networks at work. In: Zaphiris, P., Ioannou, A. (eds.) LCT 2016. LNCS, vol. 9753, pp. 663–674. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39483-1_60
65. Ziefle, M., Brauner, P., Speicher, F.: Effects of data presentation and perceptual speed on speed and accuracy in table reading for inventory control. Occup. Ergon. **12**(3), 119–129 (2015)
66. Barsky, A.J., Peekna, H.M., Borus, J.F.: Somatic symptom reporting in women and men. J. Gen. Intern. Med. **16**(4), 266–275 (2001)
67. Kroenke, K., Spitzer, R.L.: Gender differences in the reporting of physical and somatoform symptoms. Psychosom. Med. **60**(2), 150–155 (1998)
68. Ziefle, M., Schaar, A.K.: Gender differences in acceptance and attitudes towards an invasive medical stent. Electron. J. Health Inf. **6**(2), 13 (2011)

Dynamic Keypad – Digit Shuffling for Secure PIN Entry in a Virtual World

Andrew Holland and Tony Morelli(✉)

Central Michigan University, Mt Pleasant, MI 48859, USA
{holla1aa,morel1a}@cmich.edu

Abstract. As virtual reality becomes more mainstream there is a need to investigate the security of user level authentication while in the virtual world. In order for authentication methods to be useful, they must be secure, not allow for any external observers to determine the secure data being entered by the user, and also not break the immersion that the virtual world provides. Using head mounted virtual reality displays, users can interact with the world by using gaze, that is selecting objects by what the user is focusing on. This paper analyzes the security issues involved with utilizing gaze detection for secure password entry. A user study finds security issues with standard gaze based PIN input, and as a result a solution to this problem is presented. The solution shuffles the numbers on the PIN pad and finds that method to be more secure while maintaining accuracy and speed.

Keywords: Virtual reality · Security · Password

1 Introduction

Head mounted displays (HMD) for virtual reality environments allow for users to experience a virtual world without being able to see anything within their real physical environment. Typical HMDs are devices like the Oculus Rift or HTC Vive that require the device to be tethered to a computer. Other devices such as Google Cardboard or Samsung Gear VR do not require a tethered PC and are powered by a standard cell phone. While in a virtual environment it is difficult for a user to utilize a physical keyboard. In order to use a virtual keyboard within the virtual environment, standard virtual object selection is used where the user moves his head around to line up a cursor with the desired target, then holds still for a period of time until the target is selected. Although this works, it also can reveal what direction the user is looking at to other people located in the same physical space. These physical head movements could reveal how the user is interacting with a virtual keyboard and as a result secure information such as PIN codes or passwords could be captured. This paper presents a user study that reveals observers can guess the correct 5 digit PIN code within 1 digit 72% of the time, and it also provides a results of a Dynamic Keypad concept that eliminates this possibility with no significant decrease in speed or accuracy.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 102–111, 2018.
https://doi.org/10.1007/978-3-319-91581-4_8

2 Background

This paper focusses on user authentication in Virtual Reality. Virtual Reality is an environment where the user cannot see the real world. A fixed display is located in a headset very close to the user's own eyes and the user is completed immersed in the world presented on the display. Commercially available Virtual Reality headsets include the Oculus Rift and the HTC Vive. These headsets are required to be tethered to a computer that is generating the content for the display.

The Oculus Rift contains a screen resolution of 1080×1200 per eye, a 90 Hz refresh rate and a 110° field of view. The Rift has built in headphones and built in rotational tracking. In addition to sensors built into the headset, the Oculus Rift also contains optional external USB sensors (known as the Constellation sensors) for more accurate positional tracking. The tracking units track the headset as well as optional hand held controllers called the Oculus Touch. The touch controllers are tracked by the Constellation sensors and also have sensors to determine hand gestures made by the user.

The HTC Vive contains a similar architecture as the Oculus Rift. The Vive headset also contains a 1080×1200 resolution display for each eye. Unlike the Oculus Rift, the Vive contains a front facing camera that permits the user to see the real world through a video representation. The VR tracking system for the HTC Vive is a series of Vive Base stations that give a 3 dimensional view of the area and include tracking the headset and the handheld controllers. These base stations, known as the Lighthouse tracking system are larger than the sensors used in the Oculus Rift.

The PC based systems such as the Rift and Vive are a little costly and also require a decent computer to run them. For a more budget friendly VR experience devices such as Google Cardboard and Samsung Gear VR allow users to experience VR by inserting a regular smart phone into a head mounted display. These types of VR devices rely on many of the sensors, display, and processing power built into the phone. The common feature between these virtual reality devices is that the user cannot directly see the surrounding environment as the display is limited to the screen located directly in front of the user's eyes.

Virtual Reality differs from augmented reality or mixed reality in that as opposed to the user seeing only from a projected screen as in virtual reality, the user can also see the environment with his own eyes and additional information is projected on a transparent screen. The main difference is that a user can still easily see items such as keyboards to input information. Augmented devices such as Microsoft's HoloLens also allow the user to interact with the device through the use of hand gestures. These hand gestures may pose their own security risks, but it is beyond the scope of this paper. This paper focuses on data entry using head motions on virtual reality devices. The rest of the paper is organized by analyzing the current state of the art when it comes to object selection in a virtual space, then current authentication methods are analyzed, and finally a new method with some experimental results is presented.

3 Related Work

In order to properly evaluate potential methods of user level validation in virtual reality, different types of validation and object selection both inside and outside of virtual reality were reviewed. Some of the most relevant examples are listed below and they are broken down into three different categories: Interaction Techniques, Security, and Representation.

3.1 Interaction Techniques

Manipulating objects virtual using a pointing device such as a Nintendo WiiMote was common several years ago. However, the issue with the WiiMote is that it senses its position by using the controller based Infrared camera to locate LEDs in the room. In order for it to function properly, the LEDs must be set up and visible by the controller. A method (Chuah and Lok 2015) for manipulating objects with a standard phone was investigated. They were able to successfully integrate navigating in a virtual reality based environment with a standard phone by utilizing the built-in sensors on the phone.

Using a user's own hands are a method of interacting within virtual reality if the virtual reality system is equipped with sensors capable of sensing where the user's hands are located in 3D space. Although this is a common method, it still has many issues (Argelaguet and Andujar 2013) with real world target selection. That study found issues with occlusion, visibility mismatch and depth perception in stereoscopic displays. It also found that interactions in virtual reality is more physically demanding than that of normal computer interactions. The study also found pointing limitations within the human motor system.

Selection techniques for wearable virtual reality systems were evaluated in another study (Brancati et al. 2015). In this study three different types of object selection tasks were analyzed including wait to click, air tap and thumb trigger. In this study it was shown that wait to click was the most effective. In this type of interaction, the user will point with his finger at an object and hold the finger there for a period of time before the system recognizes the gesture as a desired selection action. The implementation used in this paper is based on this wait to click process, however instead of using cameras to track the position of the hand, gaze detection was used.

More target selection research (Velloso et al. 2015) analyzed the difference between gaze based selection and hand based selection in both 2D and 3D virtual environments. The results of that study show that gaze selection is not only faster but more preferred for users tasked with selecting objects in virtual reality.

Looking at how people best interact with objects in virtual reality, it was shown (Geiger et al. 2017) that people interact best in a virtual environment when additional feedback is given to them. In this case, additional hand color feedback was given in order to maximize the performance.

3.2 Security

Computer security is a very important topic and it directly translates to virtual reality. It has been shown (Das et al. 2014) that social factors can have an effect on how users perceive security behaviors. Most importantly, it was shown that the observability of security features was a key factor in socially motivated behavior change. Because of that the user study described here involves an observer phase. This helps both with the determination of the security of entering PINs in virtual reality as well as showing users how they might best interact with such an environment.

It has been shown (Fiebig et al. 2014) that a user's interaction with a standard smartphone can be determined by examining the user's face through the built in smartphone camera. Modern smartphone cameras are equipped with a high enough resolution camera such that the users actions and even reflections off of the cornea can be determined through the front facing camera. In that Fiebig user study, users were tasked with entering PIN numbers on to a smartphone while the smart phone snapped pictures of them with the front facing camera. Different users looked at the snapshots and were able to determine the PIN numbers being entered by the users.

Augmented or mixed reality systems allow the user to see with his own eyes and that may result in a different type of authentication. One approach (Roesner et al. 2014) utilized a chrome based plugin for secure web browsing authentication. In this case, the user is authenticating to a website through the augmented reality system. When a password is required, the browser will popup a unique QR code and the augmented reality system will then display the password on the user's display. As the user is already authenticated with the augmented reality device, all lookups for the password are done completely outside the computer requesting the authentication.

3.3 Representation

Determining how to design the virtual key pad for PIN based entry, design techniques were analyzed. One project (Ragan et al. 2015) found field of view and scene complexity as two factors leading to better performance. The larger the field of view, and the lower the complexity resulted in better performance. These two factors were taken into consideration when designing the keypad and its interactions.

Commercially, (Chang and Gupta 2017) a patent has been granted dealing with object manipulation in virtual reality. In this patent, users can interact in a virtual reality based inter-net browser by spinning a virtual representation of a three dimensional graphical representation of internet search results. In this example, although the user is viewing the data in three dimensions within the virtual world interactions are performed using a computer mouse and pointer very similarly to how interactions are performed on a standard computer. Although this is a novel approach, the system described in this paper utilizes a 2D PIN pad located in three dimensional space.

4 Methods for Validation

In order to properly authenticate a user in a virtual world there are several options. The user could authenticate before entering the virtual world. The user could authenticate while in the virtual world, or the user could leave the virtual world mid-session in order to perform the authentication. Each of these are discussed below.

4.1 Before Entering the World

Using this authentication method, a user would authenticate herself by utilizing a standard authentication method before putting on the virtual reality headset. The advantages of this method include the access to any existing or future standard computer based authentication techniques.

If the virtual reality headset is attached to a standard computer, the user could authenticate herself through a standard user name and password entry through the keyboard. A successful validation will allow the user to put on the headset and remain in the virtual world as the authenticated user.

If the virtual reality headset is not attached to a standard computer, but is using a mobile phone as the screen, the user also has validation options prior to putting the mobile device into the headset. The user can be validated through a text username/password as mentioned above, or the user could use any type of biometric validation methods provided by the device's operating system such as fingerprint, iris, or facial recognition.

Although using existing validation methods has its advantages, there are also some disadvantages. One issue using these methods is that it is not a seamless experience for the user. The user must perform all the validation actions, then get into the virtual environment. Many times this ends up with the user having the headset halfway on to view the computer screen while the credentials are entered. Or the user will enter the credentials onto the mobile device, but then accidentally hit the power button on the phone as it was attempted to be inserted into the headset. This results in a frustrating experience.

In addition to a frustrating experience, it also allows for credentials to be shared. For example, a user will authenticate with the system, then the headset could be passed around from person to person without requiring a new authentication thus leaving the first person logged into the system for the duration of the experience no matter how many different people use the headset.

4.2 External Mid-Session Validation

Another approach would be to use the standard validation equipment for the platform, i.e. Mobile or Computer based, in the middle of the experience. When credentials are required, the user is requested to remove the headset and enter the correct information through a keyboard or through biometric sensors on a mobile device. This type of approach is good because it is using standard authentication methods, and can be used for multiple users.

The downside of an external mid-session validation is that it breaks the immersion of the experience. If a person is in a virtual world, and then travels to a different section or a completely different virtual world, it is likely that a subsequent validation will have to take place. Taking off the headset and potentially removing a mobile device from the headset is very time consuming and not natural for a person who is in a virtual world. This type of interruption could not only break the immersion for a single user in a single session, but it could jeopardize the entire virtual reality industry as user authentication is necessary for all computing, and immersion is necessary for virtual reality. This approach makes it difficult for both of these to ever be true.

4.3 Internal Mid-Session Validation

An alternative approach is to have the user perform the authentication within the virtual environment itself. This will address the security concerns, as well as allow the user to remain immersed and preserve the virtual reality experience. The direct path to internal validation is to use a virtual keyboard that a user will select keys on using some kind of virtual pointer. The rest of this paper analyzes this topic as there are issues surrounding this type of user authentication. There are also many benefits. User Study.

4.4 Design

In order to investigate if characters entered from a virtual keyboard could be recognized from another person observing the user in the virtual world, a user study was created. The first part of the user study was to familiarize the user with virtual keyboard entry within a virtual world using two different types of keypads. In the virtual world a standard 9 digit key pad was shown on the screen. The Static Keypad was organized in 3 columns and 4 rows with the numbers appearing in sequence starting with the keys 1, 2, 3 in the top row, 4, 5, 6 in the second row, 7, 8, 9 in the third row and in the bottom row the keys were B, 0, E. The two additional buttons B (backspace) and E (enter) allowed for the participants to correct any errors and to signify that the PIN entry was complete. The Dynamic Keypad was identical to the Static Keypad, except that the location of the numerical buttons was shuffled. Participants were instructed to select a particular virtual key by staring at the desired key for 1 s. Participants in the user study were given a random 5 digit pin code before going into the virtual environment and once in the virtual environment they were tasked with entering the 5 digit code followed by selecting the E virtual key to complete the task. Each participant entered the pin code twice. Times for each character entry, total time to enter the full pin, and any errors were recorded.

After the user entered the PIN codes within the virtual environment, the user was tasked with observing two videos of a person performing the pin entry task. All participants viewed the same 2 videos and were shown each video twice. After each video was complete, users were asked to write down the PIN code being performed in the video. The first video was of a person using the Static Keypad. The second video showed a user performing the PIN entry using the Dynamic Keypad, where the digits on the key pad were shuffled and shown in a random order. Both videos were taken from the rear perspective of the User to simulate a more realistic shoulder surfing scenario. The subjects were instructed to deduce the 5 digit PIN of the subject within the video by observing the orientation of their head. Upon completion of the video observations the users were queried on their observations from the video, and what strategies they used to deduce the PINs.

4.5 Implementation

The Pin-Based virtual reality authentication system consists of a back-end server, and a front-end user interface.

Back End Server
The Windows Operation system, Apache HTTP server, MySQL database, and PHP server-side scripting (WAMP) web development platform provides several functions for the study: (1) it contains the scripts that provide verification and authentication within the system, (2) it houses the two log systems; (i) the primary log system tracks every authentication attempt and outcome as well as the user identification number, and duration of the authentication session, (ii) the secondary log system is used to track pin positions for inputs on dynamic PIN pad interfaces, and input time in all sessions, (3) maintains a database table that contains the records of all users, and user PINs which is referenced to authorize a user's authentication attempt, and send the unlock screen notification.

Front-end User Interface
The two user interfaces are composed of a static user log-in screen and a dynamic user log-in screen.

- Static Log-in Screen
  The statically created log-in screen (Fig. 1) posted within the virtual reality environment was crafted as a three by four grid with each index holding a single numeric or command value. The model was based on standard PIN pads for automatic teller machines or credit card readers and is depicted below. The log-in screen was developed using the C# programming language within the Unity IDE version 5.5.1f1.

Fig. 1. Static screen showing standard digit placement.

- Dynamic Log-in Screen
  The dynamically generated log-in screen (Fig. 2) displayed within the virtual reality environment generates a three by four grid with each index randomly assigned a numeric value of zero to nine or a command value of enter or backspace; represented

as "E" and "B" respectively, which is depicted below. The log-in screen was developed using the C# programming language within the Unity IDE version 5.5.1f1.

Fig. 2. Dynamic screen showing random digit placement

In addition to the specific PIN pads shown above, the users would be able to see the current series of numbers selected, as well as the completed pin once the E key was pressed (Fig. 3).

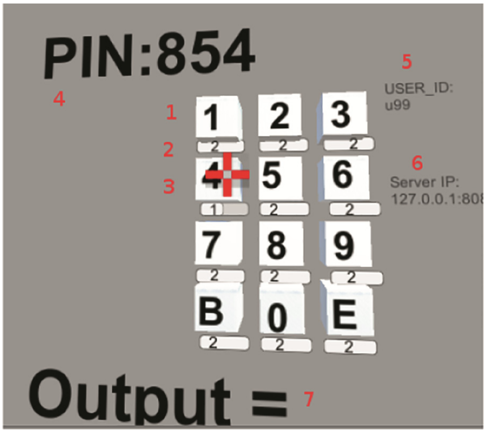


Fig. 3. User display with additional information. Numerical identifiers are as follows: (1) Static Keypad, (2) Timer counter indicating the number of seconds until that digit is selected, (3) Current selected number, (4) Current digits of the PIN, (5) User identifier, (6) IP Address of the server, and (7) Output label indicating whether or not the PIN was successfully validated.

5 Results

The user study consisted of 16 participants. It found no significant difference in speed or accuracy for users entering PIN data using the Static Keypad vs the Dynamic Keypad.

Participants observing a person entering a PIN using the Static Keypad were able to accurately determine the exact PIN 31% of the time, and with only 1 digit incorrect 41% of the time. That is 72% of the time the PIN was either exactly known or known within 1 digit. Users observing someone using the Dynamic Keypad were never able to guess the PIN within 2 digits.

The total results of the initial session showed that PIN entry speed decreased between the first 2 trials with the second PIN entry being on average 57.5% of the first (SD = 0.279). The average time to enter a complete accurate password for the second trial was 20.42 s (SD = 4.7). This is the time to enter the 5 digit PIN code and press enter with a 1 s wait time to acknowledge the user was really intending to select a digit. It is also noteworthy that none of the participants in their second attempt entered any incorrect digits. The Static Keypad was on average 2.4% faster than the Dynamic Keypad. An analysis of variance (ANOVA) shows no significant variation between the time to enter a PIN on the Static vs the Dynamic keypad, $F(1, 12) = 0.88$, $p > 0.5$.

6 Conclusion

Virtual reality presents many great possibilities. However, security in a virtual environment may have holes depending on its implementation. Object selection has restrictions based on the hardware in use. Some devices contain external controllers, others utilize cameras, and others (mixed reality) allow the user to interact with a standard keyboard. However there are many scenarios where a user is wearing a head mounted display with no access to external devices and must perform a user level authentication within the virtual space. This paper investigated the security of using head position to select a pin number on a standard keypad and found that it was possible for an external observer to determine the PIN number entered by the user by simply watching the head movements. In order to overcome that limitation, a possible solution was presented that shuffled the digits. An identified potential downside to this method could be more errors and less speed. The results did not identify any significant downside from the digit shuffling method. Due to no significant speed increase and no significant decrease in accuracy, yet an extreme increase in security, it is recommended that Dynamic Keypads be used for PIN entry in virtual environments.

References

Alhadidi, B., Arabeyat, Z., Alzyoud, F., Alkhwaldeh, A.: Cloud computing security enhancement by using mobile PIN code. JCP **11**(3), 225–231 (2016)

Argelaguet, F., Andujar, C.: A survey of 3D object selection techniques for virtual environments. Comput. Graph. **37**(3), 121–136 (2013)

Brancati, N., Caggianese, G., Frucci, M., Gallo, L., Neroni, P.: Touchless target selection techniques for wearable augmented reality systems. In: Damiani, E., Howlett, R.J., Jain, L.C., Gallo, L., De Pietro, G. (eds.) Intelligent Interactive Multimedia Systems and Services. SIST, vol. 40, pp. 1–9. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19830-9_1

Chang, S., Gupta, A.: U.S. Patent No. 9,720,562. U.S. Patent and Trademark Office, Washington, DC (2017)

Chuah, J.H., Lok, B.: Experiences in using a smartphone as a virtual reality interaction device. Int. J. Virtual Real. (IJVR) **11**(3), 25–31 (2015)

Das, S., Kim, T.H.J., Dabbish, L.A., Hong, J.I.: The effect of social influence on security sensitivity. In: Proceedings of SOUPS, vol. 14, July 2014

Fiebig, T., Krissler, J., Hänsch, R.: Security impact of high resolution smartphone cameras. In: WOOT, August 2014

Geiger, A., Bewersdorf, I., Brandenburg, E., Stark, R.: Visual feedback for grasping in virtual reality environments for an interface to instruct digital human models. In: Ahram, T., Falcão, C. (eds.) AHFE 2017. AISC, vol. 607, pp. 228–239. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60492-3_22

Huang, H., Lin, N.C., Barrett, L., Springer, D., Wang, H.C., Pomplun, M., Yu, L.F.: Analyzing visual attention via virtual environments. In: SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments, p. 8. ACM, Chicago, November 2016

Ragan, E.D., Bowman, D.A., Kopper, R., Stinson, C., Scerbo, S., McMahan, R.P.: Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. IEEE Trans. Vis. Comput. Graph. **21**(7), 794–807 (2015)

Roesner, F., Kohno, T., Molnar, D.: Security and privacy for augmented reality systems. Commun. ACM **57**(4), 88–96 (2014)

Velloso, E., Turner, J., Alexander, J., Bulling, A., Gellersen, H.: An empirical investigation of gaze selection in mid-air gestural 3D manipulation. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 315–330. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22668-2_25

VR Evaluation of Motion Sickness Solution in Automated Driving

Quinate Chioma Ihemedu-Steinke[1,2,3(✉)], Prashanth Halady[1], Gerrit Meixner[2], and Michael Weber[3]

[1] Robert Bosch GmbH, Stuttgart, Germany
{quinatechioma.ihemedu-steinke,halady.prashanth}@de.bosch.com
[2] UniTyLab, Heilbronn University, Heilbronn, Germany
gerrit.meixner@hs-heilbronn.de
[3] Institute of Media Informatics, University of Ulm, Ulm, Germany
{quinate.ihemedu-steinke,michael.weber}@uni-ulm.de

Abstract. The sensory conflict theory describes the occurrence of motion sickness caused by the discrepancy between the motion felt and the motion visually perceived. During driving, drivers monitor the environment while performing driving tasks, this enables them to get the visual perception of the motion felt. Visual cues help drivers to anticipate the direction of movement and thus, eliminate confusion, which could lead to anxiety, and thus motion sickness. Occupants of highly automated vehicles will have the luxury of performing activities such as reading or interacting with their mobile devices while the system performs the driving tasks. However, if the passenger takes his eyes off the surrounding traffic environment, sensory conflict is likely to occur. We implemented a concept in virtual reality to prevent motion sickness during automated driving based on a split screen technology. A part of the screen shows a video capture of the car surrounding in real time, while the other part is free to be used for individual applications. This additional data enables visual cues, which makes it possible to monitor the direction of movement of the vehicle. This minimizes sensory conflict and prevents motion sickness. An experiment was conducted with fourteen participants on a virtual reality automated driving simulator with an integrated motion platform. The result shows that the video streaming of the horizon presented to the passengers on a display helps them to feel comfortable and also reduced motion sickness during automated driving.

Keywords: Virtual reality · Automated driving · Motion sickness
Driving simulator · Visual cues · Sensory conflict

1 Introduction

Motion sickness (MS) is described as the conflict between the perceived motion of the body as detected by the vestibular system (VS) (an organ responsible for equilibrium) and the visual perception [1]. The brain gets different motion signals that do not correspond for the same stimulus, and thus gets confused, a phenomenon known as sensory conflict (SC) [2]. The human body's evolutionary program then comes into effect and

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 112–125, 2018.
https://doi.org/10.1007/978-3-319-91581-4_9

assumes that the body is poisoned, and responds to the restoration of health with symptoms such as nausea, headache, salivation, and in a worst case, vomiting. This is known as the "Poison theory" [3]. MS is frequent with passengers travelling by car, ship or by air, also known as travel sickness. The following three aspects characterize the critical factors for passenger-induced MS. Firstly, the discrepancy between the motion perceived by the VS and the visually perceived motion. Secondly, the loss of control of the direction of movement of the vehicle and inability to adjust to the possible direction of movement of the vehicle. Finally, MS-inducing activities [4].

One main benefit of automated driving (AD) is the time available for occupants to read or interact with mobile devices without having to deal with the traffic environment and the conventional driving tasks since the system takes over the driver's role [5]. All drivers therefore assume the role of passengers, and thus the number of passengers will increase significantly as a result of ride sharing. This also increases the number of passengers who are likely to suffer from MS while travelling by car. If the vehicle assumes the driver's role, this will increase the risk for all passengers of getting motion sick, especially those who are reading and not focusing on the surrounding traffic environment, but still perceive body motion. Activities such as reading a book or watching movies, which takes the eyes off the road are most likely to make users of AD sick [6].

Because the possibility to use the driving time effectively to perform other tasks other than driving is what makes AD attractive, MS will definitely be a big setback for the acceptance of the technology and will reduce AD benefits. Unfortunately, there are no concrete solutions on how to tackle MS effectively while travelling by road. Some sources recommend medications, acupressure wristbands which are targeted to particular groups of persons only [7]. Some sources recommend design measures to make the surrounding visible to all by avoiding tinting of rear windows, stabilizing mobile devices used while the car is in motion, encouraging all passengers to face forward, and many more [4]. These solutions still force the passengers to focus on the road most of the time, and limits in-vehicle-design, therefore not very beneficial for AD occupants.

Since most AD passengers are likely to spent most time with their smartphones and tablets, this work therefore suggests a concept evaluated in virtual reality (VR) automated vehicle simulation meant to prevent MS during AD based on a split screen technology, where a part of the screen presents the missing visual cues of the surrounding in real time, through a video capture of the surrounding traffic environment, while the other screens could be used for individual preferences (refer to Fig. 1). Passengers of AD can therefore use the available time efficiently for other activities without worrying about getting sick. Moreover, desired designs must not be eradicated in order to avoid SC (rear windows may remain tinted if so desired).

Fig. 1. Tablet with split screen shows a text and the video capture of the traffic road during.

2 Related Works

There is no limit to what self-driving cars can do when compared to a human driver with technology innovation advancing in full speed. The advantages of AD if effectively implemented and inaugurated knows no limit e.g. traffic safety, reduced traffic congestion, free and safe time to work while the car is in motion, comfort and unlimited mobility for the elderly. The vehicle will become a safe, social and working platform to hold meetings, talk on the phone with hands off the steering or relax during congested traffic situations. However, concern about MS affecting occupants during AD is rising despite numerous studies carried out on this topic for decades. A study carried out by the University of Michigan showed that by relieving the driver's role, around 30% of occupants of automated vehicles would carry out activities that could induce MS. Activities such as reading, working, watching movies and writing emails are critical factors that could lead to MS due to the opposing information of "movement" in the VS and the visual signal through focusing on a resting point [6]. Though it is not evident what really causes MS, it could be described as the presence of mental and physical discomfort (e.g. nausea, increased salivation and disorientation) due to the discrepancy between the perceived motion or missing motion and the actual motion felt [8, 9].

The VS, located in the inner ear and responsible for the sense of balance and spatial orientation is composed of the semi-circular canals. Each canal detects one of the following head movements; nodding up and down, shaking side by side or moving left and right. A fluid called endolymph moves through the canals when the head is rotated. The canals are connected by a component known as Ampulla which contains tiny hair cells with nerve impulses. When humans move, the fluid moves in the direction of the movement. The tiny hairs detect the movement of the fluid and sends nerve impulses to the brain, which helps for orientation. When the VS transmits conflicting signals from other sensory signals, this results to sensory conflict and explains why travelers get motion sick, most especially when passengers are reading while the car is in motion. In this case, the VS detects car motions while the eyes see only texts. The brain receives conflicting signals and this results to SC, which in turn causes discomforts.

Passengers are usually more affected than the driver because the driver can anticipate the next move and therefore knows what to expect and can readily adapt [10]. Passengers who are not always looking outside the car in the direction of movement, might be ignorant of the speed, direction of movement, and might not anticipate a bad curve. In order to adapt to the next move, they have to look out the window so that the visual and VS can detect and agree on the sense of motion. It is therefore recommended that passengers look at the horizon so that the eyes can detect the motion felts by the body. This will help to eliminate SC, though susceptibility to MS differs from person to person.

Many studies have been carried out and still ongoing on how to avoid MS during AD. Some studies recommend a constant speed to reduce MS. Other sources suggest the interior vehicle design modification such as avoiding rear-sitting arrangement and tinted glasses in order to make the horizon visible [8]. Another source recommends working devices such as displays or tablets to be positioned in such a way that the horizon is still visible to the eyes for possible visual cues [10]. AD will allow for more connectivity and interaction with mobile devices, it is therefore expected that passengers will most likely interact more with devices such as tablets and smartphones. Therefore, designers are recommended to focus on the stabilization of these devices for a steady information presentation and also consider user interface designs mentioned above that reduce MS [4]. Finally, another study investigated two different approaches to reduce MS while watching movies on onboard displays. The first approach used vertical stripes as a surrounding image which is perceived as a background, which induced circular vection to a certain degree. In the second approach, movies were displayed as if the plane of the screen rotated on a vertical axis, and circular vection was induced as in the first approach. And because circular vection reduces SC, MS was reduced [11]. While considering various design approaches to combat MS during AD, the approaches have to be validated for safety in order to avoid injury or risking the lives of the occupants [8]. Though all the approaches and solutions mentioned above could help to avoid MS during AD, they place some constraints which could reduce the benefits of AD.

3 Virtual Automated Vehicle Implementation

In order to investigate the impact of the visual cues presented to passengers during AD, an automated vehicle prototype was developed in VR environment using Unity 3D engine and Oculus Rift consumer version 1 (CV1) [12, 13]. The implemented solution is a virtual automated vehicle with a simple but effective takeover manoeuver which enables both manual and AD on a typical German highway with few traffic features. The simulation of a highly detailed real-world traffic elements and integration of VR technology is meant to improve the realism of the implemented traffic scenario [14]. Figures 2 and 3 show the road network layout and the simulated three-lane German highway respectively. A Tesla-like vehicle model was initially developed and textured.

The vehicle interior was designed based on a previous Bosch automated vehicle human-machine-interaction concept [15].

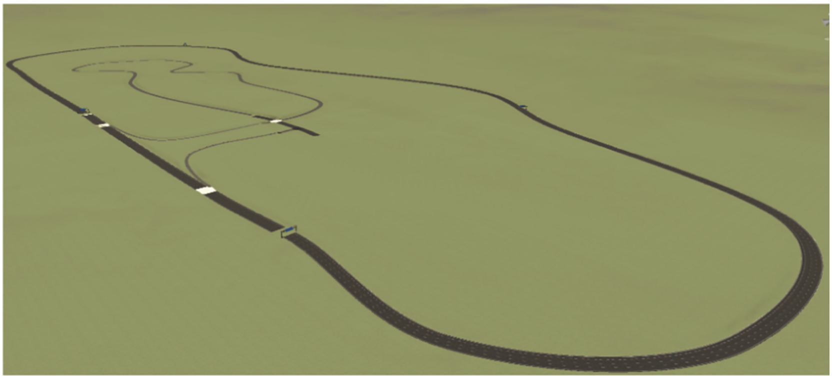


Fig. 2. Layout of the implemented road network.

Fig. 3. The virtual 3-lane typical German highway.

3.1 Development Tools

VR applications require high computing intensity due to the increased total number of pixels, high refresh rate, and because 3D scene has to be rendered separately for each eye at the same time. Therefore, powerful components are needed to accomplish these tasks. This project was developed with a VR-Ready computer with the processor Intel Core i7 5960X with 8 cores and 16 threads, as well as clocking up to 3.5 GHz. The NVidia Titan X graphic card used offers a variety of functions e.g. simultaneous multi-projection, which is automatically rendered specifically for VR for a higher performance and a more detailed virtual environment to increase realism [16].

The Oculus CV1 used for this work offers a resolution of 2160 × 1200 pixels for both eyes, and a refresh rate of 90 Hz. The higher resolution as well as the refresh rate helps to achieve a higher level of immersion in the virtual environment and also reduces the occurrence of simulation sickness during user evaluation. Oculus CV1 was chosen because of ergonomic preferences (light-weight and comfortable) and the high-performing audio. It is also well supported by Unity 3D for a simple and fast integration.

Logitech electric steering wheels and pedals are used to reproduce the steering, accelerating and braking controls of the virtual vehicle, which is then passed on to the computing system. In this project, the Logitech G29 steering wheels, pedals and shifters were used. The red button on the steering wheel was implemented to help the drivers choose a profile, and also scroll through the texts while reading. Logitech offers a Unity plug-in as well that facilitates integration.

Atomic A3 of the Atomic Motion Systems (AMS) was integrated to the VR driving simulator as a cost-effective way to transfer motion from the virtual world to the real world. It offers the possibility to move the drivers on two axes, each with a 27° angle of incidence and a speed of approximately 72° per second. AMS provides native Unity support for Atomic A3. Simphynity software with which it is possible to run the platform, detects motions of the implemented vehicle dynamics and reproduces these forces in form of motion. The Atomic A3 2 degree of freedom motion simulator was integrated to the system in order to simulate the feeling of a moving vehicle (refer to Fig. 4).

Fig. 4. Atomic A3 from Atomic Motion Systems.

4 Experiment

The study separated participants into two randomized groups with each group having equal number of male and female participants: a control group which only got displayed texts on the tablet, and the experimental group which got both texts and video streaming of the traffic environment on a split screen in real time (refer to Figs. 7 and 8). In order to collect and measure data, two questionnaires were used. The first questionnaire consisting of thirteen questions, was established to collect socio-demographic information, previous driving experience, and VR experience. The second questionnaire is the simulation sickness questionnaire (SSQ), which consists of 16 questions and was developed by Kennedy and colleagues in 1993 (refer to Fig. 5) [17].

		None	Slight	Moderate	Severe
1	General discomfort				
2	Fatigue				
3	Headache				
4	Eye strain				
5	Difficulty focusing				
6	Salivation increasing				
7	Sweating				
8	Nausea				
9	Difficulty concentrating				
10	« Fullness of the head »				
11	Blurred vision				
12	Dizziness with eyes open				
13	Dizziness with eyes closed				
14	*Vertigo				
15	**Stomach awareness				
16	Burping				

Fig. 5. Simulation sickness questionnaire. Original version by Kennedy and colleagues [17].

4.1 Participants

The experiment comprised of fourteen participants (N = 14), eight males (Mage = 38.34, SD = 10.28) and six females (Mage = 34.83, SD = 5.42). The participants were aged between 28 to 55 years old. Two participants had previous cases of MS, seven were never affected while five were not sure if they suffered from MS before. Ten out of the fourteen participants had previous experience with driving simulators (refer to Fig. 6). More than 85% of the participants are frequent drivers. Participation was voluntary and no compensation was awarded to the participants. All participants have a university degree.

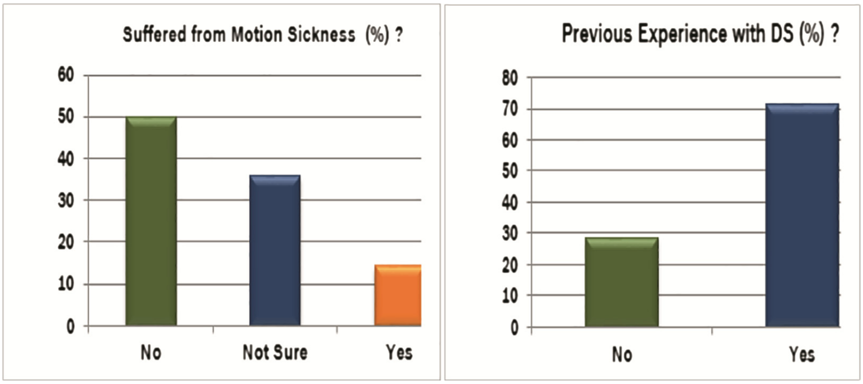


Fig. 6. Most users never suffered from motion sickness before (left) and most had previous experience with driving simulators (right).

4.2 Procedure

The experimental group was presented with a multiple screen tablet, which displayed both texts and video streaming of the horizon in real time, while the control group was presented with a tablet with texts only to be read out loud for ten minutes during AD. Firstly, each participant completed the pre-questionnaire which mostly consisted of socio-demographical questions, previous experience with driving simulators, driving experience, MS history and previous experience with VR headset. Because most participants were working prior to the experiment, and most probably experienced some symptoms such as fatigue, the SSQ was completed before and after the experiment in order to effectively compare the initial and current state of wellbeing of the participants.

Secondly, instructions were giving on how to operate the system and navigate freely in the virtual environment. This was followed by a two-minute test drive in order to enable the users to get acquainted with the system and eliminate anxiety during the experiment. The experiment proper starts with the system requesting the users to drive manually for two minutes, after which they can perform a takeover and hand over control to the system. The system then controls the vehicle for approximately ten minutes. In the autonomous mode, a tablet appears with captivating African fairy tales which all participants read out for ten minutes. While reading, the participants were not allowed to look at the horizon or get any visual cue of the surrounding traffic environment. They could look around the vehicle inner design or stay focused on the reading assignment. Motion cues was provided with the motion simulator in order to give the feeling of driving on a real road. Participants of the multiple screen group were presented with the video capture of the surrounding on the lower section of the tablet (see Figs. 1 and 7) while participants in the control group got only texts displayed (see Fig. 8). Finally, the experiment ends with the participant completing the SSQ for the second time. Each test slot lasted between thirty and forty minutes depending on how fast the users completed the questionnaires. None of the participants dropped out nor stopped the experiment before the simulation ended.

Fig. 7. User of the experimental group reading texts while viewing the displayed horizon.

Fig. 8. Only texts presented to the control group during automated driving.

5 Results

This section presents the results of the experiment while displaying the total score of typical symptoms of MS based on the SSQ score for both groups. The result of the experiment revealed that participants with the split screen suffered less from those symptoms peculiar to MS, when compared to the control group who were only presented texts to read and no visual cues of the horizon during AD.

To learn how the SSQ score of each symptom is calculated, refer to [3, 17]. The score distribution for the symptom "General discomfort" for the experimental group was

9.5% before exposure as compared to only 4.7% after exposure. Meanwhile, the score distribution for the control group was 14.3% for "General discomfort, as compared to 9.5% before exposure (refer to Fig. 9). Therefore, the control group suffered more general discomfort. Most participants of the experimental group described the experience as relaxing while the control complained it was strenuous and boring. Likewise, the severity level for the symptom headache did not change for the experimental group, unlike the control group which suffered severe headache after exposure. The score distribution of the control group for the symptom "Headache" was registered as 4.7% before the experiment, as compared to 10% after exposure. No headache was recorded before and after exposure for the experimental group (see Fig. 10).

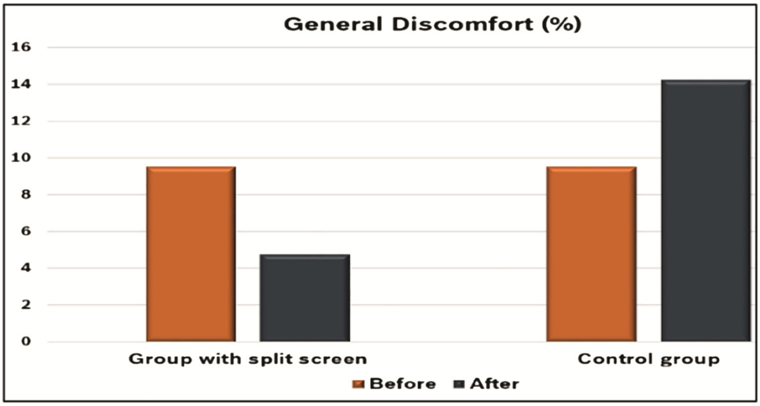


Fig. 9. Before and after comparison of general discomfort maximum score for both groups.

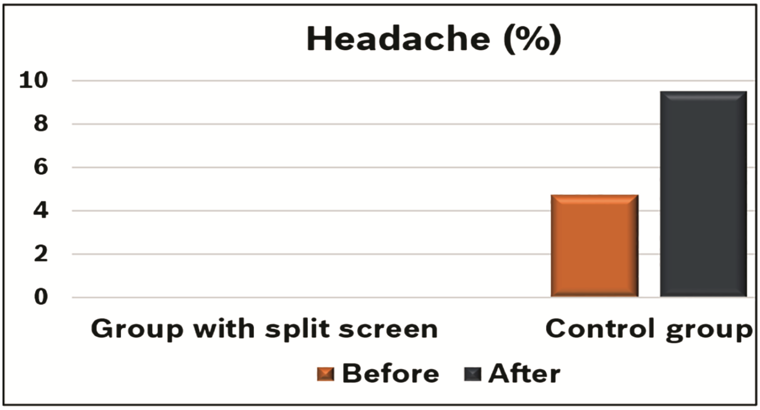


Fig. 10. Experimental group suffered no headache. Control group suffered severe headache.

Figures 11, 12, 13 and 14 show the results of the symptoms salivation increasing, stomach awareness, nausea and eye strain for both groups. Figure 11 shows that 9.5% maximum score was recorded for the experimental group before and after exposure for the symptom "Salivation increasing". The control group suffered more with a 4.7% before as against 9.5% after the simulation. Meanwhile, no prior stomach awareness discomfort was recorded for both groups. Very high stomach awareness discomfort was recorded after exposure for the control group as compared to the experimental group (Fig. 12).

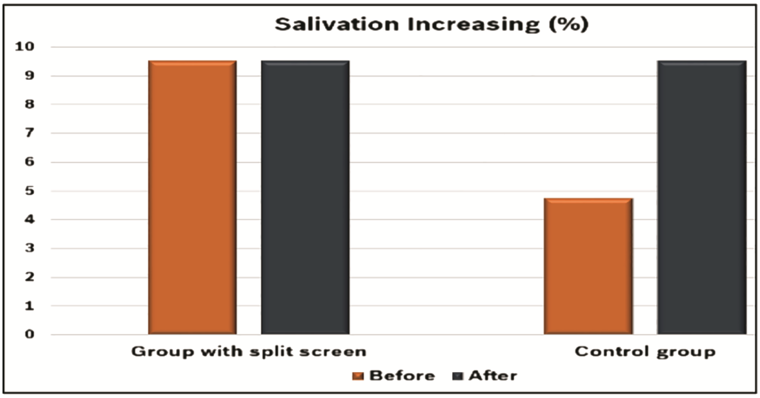


Fig. 11. Before and after results of salivation increasing for both groups.

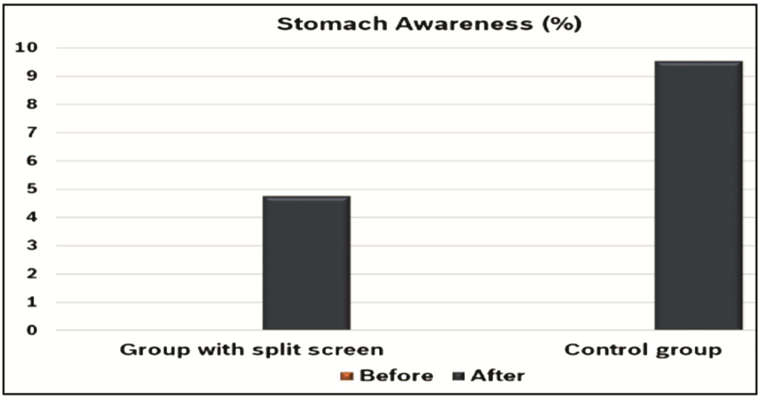


Fig. 12. Before and after results of stomach awareness for both groups.

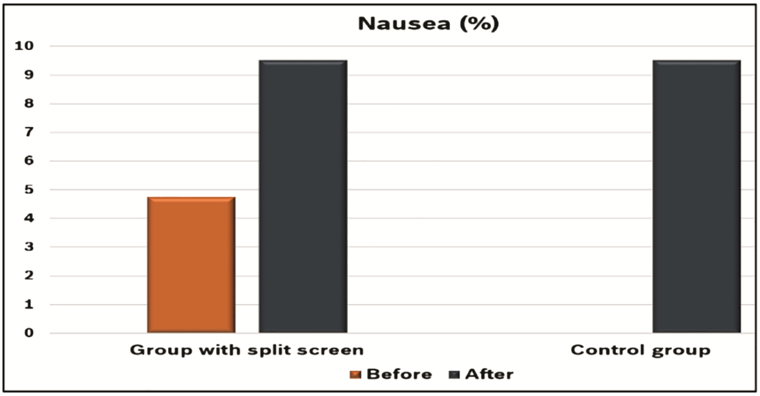


Fig. 13. Before and after result of the symptom nausea for both groups.

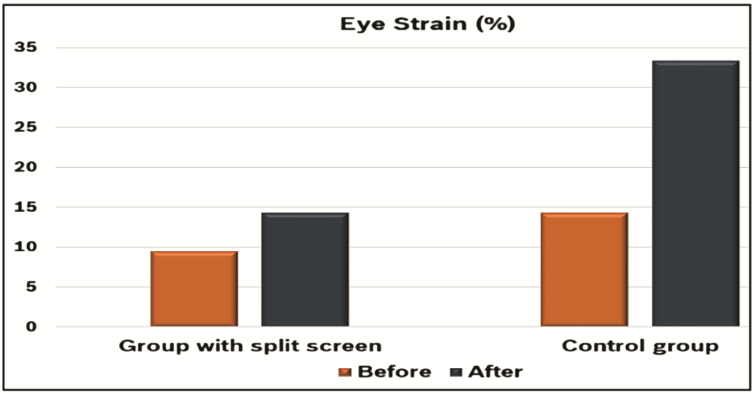


Fig. 14. Distribution score for eye strain for both groups.

Meanwhile, other symptoms which are not graphically represented here e.g. vertigo recorded a 0% before and 4.8% after exposure for both groups. Burping also recorded a 4.8% score distribution before and after exposure for both groups. The maximum score recorded for the symptom "sweating" for both groups remained the same before and after exposure. Symptoms such as "difficulty concentrating" showed a slight advantage for the experimental group when compared to the control group. On the other hand, the experimental group (before = 38.1%, after = 38.1%) was more affected by the symptom "fullness of the head" than the control group (before = 28.6%, after = 19%). Finally, the symptom "fatigue" was greatly reduced for the experimental group after the exposure (before = 43%, after = 29%) as compared to the control which was only slightly reduced (before = 29%, after = 24%). Though eye strain is peculiar to simulation sickness [18], the result was also considered as important because the experiment was carried out in a VR environment (refer to Fig. 14).

6 Discussion and Future Work

This work evaluated whether using a multiple screen display which streams the traffic environment in real time, could help reduce MS during AD. We implemented a VR solution for MS using Oculus rift CV1. The implemented solution is based on the projection of a video capture of the surrounding vehicle environment on a multiple screen device such as a tablet or vehicle onboard display, which enables the occupant to view the horizon while performing other activities during AD. The motion simulator was integrated to the system in order to simulate the feeling of a moving vehicle. An experiment was conducted with fourteen participants and the results show that occupants of automated vehicles require visual cues of the horizon while reading or carrying out activities which take their eyes off the road while being driven. This helps to reduce the discrepancy between the motion they feel and the motion they do not see while reading.

Typical symptoms of MS such as stomach awareness, headache and nausea were mostly higher for the control group which gives a cause for alarm for AD passengers. The occurrence of MS could greatly affect the acceptance of AD because the fear of MS could make drivers to avoid activities that could make them sick, or just focus on the road in order to avoid sickness. This will hinder the effective usage of the time available and the luxury of being driven, thereby greatly reducing the benefits of self-driving cars [8]. When the driver takes up the role of the passenger during AD, they should not be under any obligation to focus on the road at all times because of the fear of getting motion sick, this could greatly reduce the benefits of automated vehicles. However, it is important to get the visual perception of the motion that the body feels in order to avoid the brain getting mismatched motion signals. When this happens, the result is sensory conflict between the motion visually perceived (the eyes tell about the texts being read by the driver) and the motion the body perceives (the vestibular system tells another story: the person is moving). This could result in most people to MS.

Occupants of automated vehicles are most likely to indulge in activities that will warrant them to take their eyes off the road such as reading, and this might result to MS. This could be avoided by generating moving image signals and optical representation of the vehicle surrounding by a vehicle camera, and presenting these video data to the passenger through a display. The display provides multiple screen, one for the real time images of the moving vehicle environment, and the other screens for work or just interaction. This will enable passengers to fully enjoy the luxury of being driven and concentrate on other tasks not related to driving. Though the implemented VR solution shows an improvement and a potential solution to MS in automated vehicles, a thorough study with more participants on a real car should further solidify and demonstrate this solution.

Future work should focus on the implementation of a real physical prototype to be integrated and evaluated in a real car. A user experience evaluation with a significant number of participants is also relevant in order to gather sufficient data for an effective evaluation of the solution. The acquired results should help in facilitating the development and improvement of the final product.

AD will offer numerous benefits such as safety, reduced traffic congestion, flexible connectivity, comfort, and mobility to many without driver's license and the elderly. It is therefore important to implement solutions that will eliminate potential limitations.

MS is one limitation that should not be taken for granted during AD because this could drastically reduce the benefits, and it is a concern to all stakeholders [19].

Acknowledgments. The authors wish to thank Vincent Musch for his contribution during the virtual prototype development and the user evaluation experience. Our appreciation also goes to Martina Deininger and Kerstin Mayr for their contributions. This motion sickness solution in automated vehicle has been registered as a patent in Germany by Robert Bosch GmbH.

References

1. Reason, J.T., Brand, J.J.: Motion Sickness. Academic Press, London (1975)
2. Oman, C.M.: Motion sickness: a synthesis and evaluation of the sensory conflict theory. Can. J. Physiol. Pharmacol. **68**(2), 294–303 (1990)
3. Ihemedu-Steinke, Q.C., et al.: Simulation sickness related to virtual reality driving simulation. In: Human Computer Interaction International 2017, Vancouver (2017)
4. Diels, C., Bos, J.E.: User interface considerations to prevent self-driving carsickness. In: AutomotiveUI 2015 Adjunct, Nottingham (2015)
5. Hars, A.: Autonomous cars: the next revolution looms. In: Inventivio Innovation Briefs 2010-01, Nuremberg (2010)
6. Sivak, M., Schoettle, B.: Motion sickness in self-driving vehicles. Michigan (2015)
7. Miksch, M., et al.: Motion sickness prevention system (MSPS) – reading between the lines. In: AutomotiveUI 2016, Michigan (2016)
8. Green, P.: Motion Sickness and Concerns for Self-Driving Vehicles: A Literature Review. Transportation Research Institute (UMTRI), The University of Michigan, Michigan (2016)
9. Onuki, Y., Ono, S., Kumazawa, I.: Air cushion: a pilot study of the passive technique to mitigate simulator sickness by responding to vection. In: IEEE Virtual Reality 2017, Los Angeles, CA (2017)
10. Kuipera, O.X., Bosa, J.E., Cyriel, D.: Looking forward: in-vehicle auxiliary display positioning affects carsickness. Appl. Ergon. **68**, 169–175 (2018)
11. Morimoto, A., Isu, N., Okumura, T., Araki, Y., Kawa, A., Masui, F.: Image rendering for reducing carsickness in watching onboard video display. In: Digest of Technical Papers - International Conference on Consumer Electronics, Las Vegas (2008)
12. Unity 3D: Unity 5.0 (2017). https://unity3d.com/de/unity/whats-new/unity-5.0
13. Oculus Rift. https://www.oculus.com/rift/#oui-csl-rift-games=mages-tale. Zugriff am 2018
14. Scholz, M., Orozco Idrobo, A.M.: Supporting the implementation of driving simulator environments through established GIS approaches by extending the geospatial data abstraction library with OpenDrive. In: Driving Simulation Conference 2017, Stuttgart (2017)
15. Bosch Mobility Solutions, Bosch (2015). https://www.youtube.com/watch?v=zPMwes6HoOA
16. Musch, V.: Virtual Reality Fahrsimulation mit autonomen Fahren - Bericht zur Praxisphase, Leonberg (2017)
17. Kennedy, R., et al.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**, 203–220 (1993)
18. Aykent, B., Merienne, F., Guillet, C., Paillot, D., Kemeny, A.: Motion sickness evaluation and comparison for a static driving simulator and a dynamic driving simulator. Proc. IMechE, Part D: J. Automob. Eng. **228**, 818–839 (2014)
19. Diels, C.: Will autonomous vehicles make us sick? In: Contemporary Ergonomics and Human Factors (2014)

Enactive Steering of an Experiential Model of the Atmosphere

Brandon Mechtley(✉), Christopher Roberts, Julian Stein, Benjamin Nandin, and Xin Wei Sha

Synthesis, School of Arts, Media, and Engineering, Tempe, AZ, USA
{bmechtley,jestein,benjamin.nandin,xinwei.sha}@asu.edu,
croberts008@regis.edu
http://synthesis.ame.asu.edu/

Abstract. We present a stream of research on *Experiential Complex Systems* which aims to incorporate responsive, experiential media systems, i.e. interactive, multimodal media environments capable of responding to sensed activity at perceptual rates, into the toolbox of computational science practitioners. Drawing on enactivist, embodied approaches to design, we suggest that these responsive, experiential media systems, driven by models of complex system dynamics, can help provide an *experiential, enactive* mode of scientific computing in the form of perceptually instantaneous, seamless iterations of hypothesis generation and immersive gestural shaping of *dense* simulations when used together with existing high performance computing implementations and analytical tools. As a first study of such a system, we present EMA, an Experiential Model of the Atmosphere, a responsive media environment that uses immersive projection, spatialized audio, and infrared-filtered optical sensing to allow participants to interactively steer a computational model of cloud physics, exploring the necessary conditions for different atmospheric processes and phenomena through the movement and presence of their bodies and objects in the lab space.

Keywords: Responsive media · Experiential media
Enactive interfaces · Computational steering · Scientific computing

1 Introduction

Our Experiential Model of the Atmosphere (EMA) is part of a larger research stream on creating computational platforms for integrated, gestural interaction with complex models via multi-modal interfaces that will allow for fluid human-in-the-loop control of computerized scenarios.[1] The main challenge to developing such a system is handling new densities of data that approach a continuous distribution. Our strategy is that an effectively continuous dynamical systems

[1] https://vimeo.com/synthesiscenter/slsa2 shows a responsive environment adapting itself ***ad hoc*** to its inhabitants' social activities all day long.

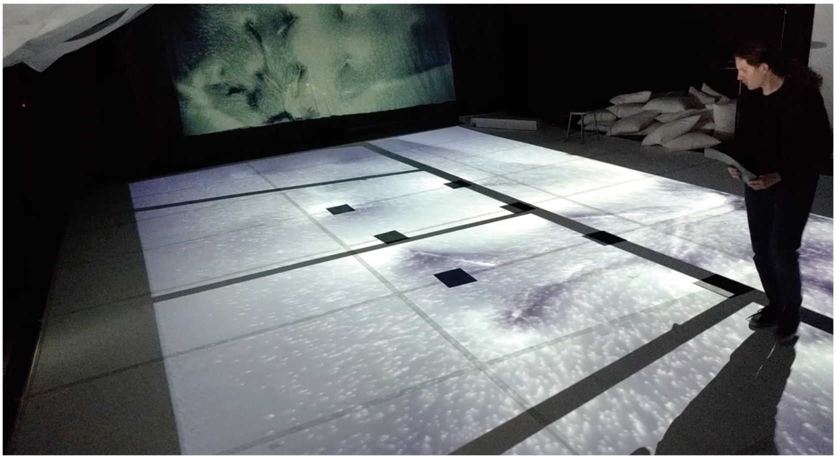
J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 126–144, 2018.
https://doi.org/10.1007/978-3-319-91581-4_10

approach can provide principles for designing a system able to evolve in real-time to non-discrete, multi-user gestural control of rich experiential scenarios that tap embodied, human experience (Varela [1], Dourish [2], Sha [3–5], Ingalls [6]). Thus, we seek to develop computational paradigms that will allow designers and users to leverage the full potential of the increasing density of sensors and computational media in everyday situations by providing a wholly experiential means of controlling and interacting with dense sensors and media. One way to scaffold our technical and design imaginary is to use the model of the swimming pool in place of the model of a graph. How does the water coordinate its activity with the activity of its inhabitants and the wind blowing across its surface? Many forces modulate its movement and condition. Some forces are due to people swimming through the water, pushed by and pushing the liquid surrounding them. Others are due to the waves on its surface, or currents distinguished by momentum, or in the case of the deep ocean, salinity and temperature. Still others are due to the wind which acts continuously across a continuous surface – the continuously extended interface between the air and the water. Note that whereas we may regard a rock thrown into the water or a swimmer as a compact, point-like source of motive force, aggregates of entities or even more essentially, extended continuous fields do not fit this model of an atomic agent. Dyadic (1-1) relational interaction is a small, sparse subset of much richer fields of experiential dynamics. Thus, we seek a more ample way to conceive engagement between different fields of media in a responsive environment.

Our method is to look for computational adaptations of continuous (e.g. differential geometric or topological) models to the scientific analysis of dense, heterogeneous environments like weather systems and urban spaces. These continuous models complement discrete models (e.g., discrete graphs) of procedural computation processes. We adopt techniques from signal processing and computer science that are also shared with machine perception, fault tolerant systems, or autonomous systems but we do so with the distinctive intent to keep human-in-the-loop control of the experience that can give designers computational paradigms leveraging *collective, embodied* experience [1,2,7–10]. The three classes of continuous models we investigate are (1) homogeneous generalized computational physics of materials, (2) continuous evolution of metaphorical states, and (3) heterogeneous atmospheric models, such as models that mix for example agent-based models of urban dynamics, models of geophysics, or rule-based systems that model interventions by large scale sociopolitical institutions, that condition the lattice models of atmosphere itself.

To calibrate these models of dense experiential systems against real-world use scenarios, we have leveraged substantial experience and collaborator expertise: live action in performative environments (dance, musical or theatrical performance, games) [6,11–14], movement/gesture tracking in everyday and rehabilitative contexts [15,16], and experiential atmospheric models based on work by experienced atmospheric scientists [17–20].

We first lay the framework for embodied, enactivist [1,2,7,21] approaches to the design of computer interfaces, and more generally of responsive environments augmented by computation. In that context we will define what we mean by



(a) A researcher studies flow between multiple low pressure regions by placing sheets of paper, whose silhouette is mapped to a constant reduction in pressure. Particles flow continuously with the velocity field, depicted by a dense pseudocolor plot. Select tracer particles are sonified as spatialized voices.

(b) Two researchers study changes to a simulated vortex sheet by narrowing a slit between two poles. The angle of air velocity is mapped to hue, with a constant external source of wind coming from upstage. Velocity magnitude orthogonal to the wind is sonified by averaging the velocity field over a coarse grid and spatializing the sound according to the grid cell centers.

Fig. 1. Realtime corporeal interaction with dense, high-dimensional, GPU-accelerated simulation of atmospheric dynamics. On the order of 100 sound processes provide spatialized sonic textures with a palpable landscape for enactive, embodied engagement. B. Mechtley, C. Rawls, Synthesis 2018.

realtime, multimodal, whole-body, gestural and multi-person engagement with an immersive responsive environment (Fig. 1).

An important motivation and context for our work is the focus on *whole experience*, in the senses of James [22], contemporary phenomenological work on experience (Gendlin [23], Casey [24–26], Morris [10], Petitmengin [27]), and movement-based experience (Sheets-Johnstone [28–30]). Under these approaches, experience cannot be decomposed into a finite number of perceptual or functional component dimensions and reassembled in some linear superposition of independent features. Senses of rhythm and of mathematical pattern are examples of such *apperceptions*. Despite this irreducibility of experience, this non-decomposability of experience into "independent" sensory dimensions, there are useful means of ascertaining accounts of experience that can be shared objectively across instances: notably methodological and experimental approaches by Petitmengin [31], Sha [32,33], and Bregman [34]. Indeed as Bregman commented in his keynote on auditory scene analysis on subjectivity "versus" objectivity:

> At this point, I want to interject a few words about subjectivity and objectivity in psychological research. The personal experience of the researcher has not fared well as acceptable data for scientific psychology. Since the failure of Titchener's Introspectionism, a very biased form of report of one's experience, in the early twentieth century, and the rise of Behaviourism to replace it, scientific psychology hasÊ harboured a deep suspicion of the experience of the researcher as an acceptable tool in research.
>
> You would think that the study of perception would be exempt from this suspicion, since the subject matter of the psychology of perception is supposed to be about how a person's experience is derived from sensory input. Instead, academic psychology, in its behaviouristic zeal, redefined perception as the ability to respond differently to different stimuli, bringing it into the behaviourist framework. We may be doing research nowadays on cognitive processes, but the research methods are, on the whole, still restricted to behaviouristic ones. Since it was a perceptual experience of my own (the rapid sequence of unrelated sounds) that set me off on a 40-year period of study of perceptual organization, I have always questioned the wisdom of this restriction.. . .
>
> Sometimes we have used both types of measures, subjective rating scales and measures of accuracy, either in the same experiment or in a pair of related experiments. The two measures have given similar results, but the subjective rating scales have been more sensitive. I think the reason for their superiority is that they are a more direct measure of the experience, whereas turning one's experience into the ability to form a discrimination between sounds brings in many other psychological processes that are involved in comparison and decision making.
>
> As a result of my belief in experience as an important part of Psychology, I'm going to try to describe some of my research on auditory perception,

> but I won't give any data. Instead, I'm going to support my arguments with audio demonstrations to the extent that time permits. [34]

Experimental platforms scaffolding such whole experiences – *experiential systems* and *responsive environments* – have been built by Sundaram [35,36], Wei [37], and others (see survey on responsive environments by Bullivant [38]). By *embodiment* we mean sense-making which is conditioned on one's corporeal engagement with the material world. By *material* we mean the union of physical, energetic, social, and affective fields (Sha [4], Massumi [39]).

Francisco Varela, Evan Thompson, and others introduced the notion of *enactive* experience to describe how we progressively construct our sense of, concepts and know-how about the material world through engagement and empirical experience: "We propose the term enactive to emphasize the growing conviction that cognition is not the representation of a pregiven world by a pregiven mind but is rather the enactment of a world and a mind on the basis of a history of the variety of actions that a being in the world performs" (Varela [1]: 9). We extend that cognitivist sense of *enaction* to a more thoroughly processualist one of how subjects, organisms (Longo and Matevil [40]), technical ensembles (Simondon [41]), more generally any individuals and their environment co-construct each other (Simondon [42,43], Sha [4]) via structural interaction (Maturana [44,45]).

We now turn to more specific qualities of experiential systems: that they are immersive, multimodal, realtime, and multi-person. Immersivity can be more precisely framed in the phenomenological distinction between acting, being, sensing in the world without any reflection – *thrownness* (*geworfenheit*), versus the state of being reflexively aware of one's stance with respect to the world (called "defamiliarization" or *Verfremdungseffekt* in some technical contexts [46]). In this more precise sense, being immersed into a situation is independent of the sensory modalities that are being most exercised. One can be immersed in reading a book on one hand, and on the other, be largely "clinically" disengaged even in a full-body, physical interaction.

Our experiential systems are designed multimodally, that is, the software framework for our experiential systems, *SC*, is designed for integrated, gestural interaction with complex models via multi-modal interfaces that allow fluid human-in-the-loop control of densities of data that approach a continuous distribution. The system evolves in real-time to non-discrete, multi-user gestural control of rich experiential scenarios, which tap embodied, human experience (Varela [1], Dourish [2], Sha [3–5], Ingalls [6,47]). Thus, we leverage the full potential of the increasing density of sensors and computational media in everyday situations by providing a wholly experiential means of controlling and interacting with dense models.

It is essential to underline that our interpretation of multimodality is sharply different from the standard sense of the simple union of computer-synthesized fields of light and sound (or other media). Rather than limit the design of the user experience to a small number of those digital synthetic media modalities we start with the full sensorium given in *physical* experience – everything that one can feel including floor and wall treatments, furniture, clothing, physical props, analog

HVAC, illumination, and acoustics – and carefully modulate certain modalities: e.g. sense of pressure or heat on the shoulder, or a field of sound and vibration from underfoot, or the thrum of video or structured light on the skin and the floor. In other words, instead of "zeroing out" the world and presenting only a few synthetic bits of media against a perceptual void, we leverage the affordances of analog furniture, media, objects, and other, co-participant bodies (Fig. 2).

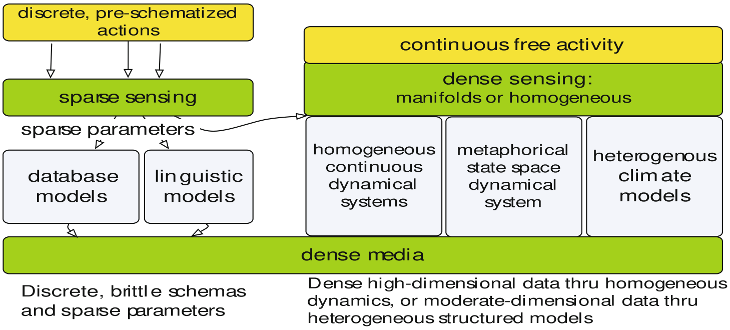


Fig. 2. Layered activity tracking and computational media processing for experiential environments.

Finally, our responsive environments are all designed for multi-person use, which requires a different sort of design than extrapolating from the design of "single-user interaction" where a single user is seated in front of a screen with keyboard and mouse WIMP interfaces that can only be controlled by one person at a time. This is a concrete setting for designing for human-human and human-system interaction based on *ensemble* experience and on *ensemble* activity. Concretely, ensemble interaction concerns situations where there are three or more human participants, so that we do not fall back on social conventions encoded in dyadic interaction. Also, this sidesteps human-machine interaction design that is implicitly predicated on single-user WIMP interface design including WWW document interfaces and most non-game "applications" whether on mobile or desktop computers. A simple example of n-person engagement ($n \geq 3$) is walking in a circle to stir up the atmosphere or the ocean model to form a large vortex (Fig. 3).

In parallel research, we have collaborated with experts in ensemble experience design from the areas of performing arts, exhibition design, and urban and landscape architecture, though those experts would not use the same term, and have broader sets of concerns than human-computer interaction. (Indeed those broader concerns provide useful ground-truth and use-scenario checks on the expressive power and robustness of our engineered systems.)

(a) A trio forms warm clouds in a simulation mapping optical flow to an increase in water vapor and temperature. On the scrim, the condensed liquid water mixing ratio is sonified by segmenting the field into eight vertical strips and mapping the field averages and differences to amplitudes and filter properties of a sample-based audio instrument.

(b) A group walks in a ring around a pivot to simulate a hurricane.

Fig. 3. Ensembles ($n \geq 3$) steering a realtime simulation by coordinated whole-body interaction. B. Mechtley, M. Patzem, and C. Rawls. Synthesis 2018.

1.1 Steerable Scientific Simulations

Creative experimental scientific work relies on constructing fresh instruments of observation *in tandem with* fresh theoretical interpretations of freshly observed phenomena. We call this on-the-fly co-construction of theory, instrumentation and observation, which is characteristic of creative work in science as well as other disciplines, *abductive* method (Morris [48], Peirce [49], Psillos [50]).

Some computational science applications have also adopted human-in-the-loop modulation of parameters through the use of computational steering. In computational steering of simulations, investigators change parameters of computational models on the fly and immediately (or as close to immediately as possible) receive feedback on the effect, in parallel with the execution of the simulation. In practice, computational steering allows investigators to quickly explore alternative paths of evolution of system state, such as through introducing exogenous changes to boundary conditions or simulation parameters.

Computational steering has been applied to the real-time control of scientific simulations, such as fluid dynamics [51] in general, air safety [52], flood management [53,54], particle physics [55], astrophysics [56], and cardiology [57], and several frameworks have been created for integrating the methodology into new and existing simulations deployed on high-performance computing platforms, including SCIRun [58], RealityGrid [59], and WorkWays [60].

We conceptually extend the notion of computational steering to real-time human-in-the-loop modulation of any computationally-modulated environment where the results are immediately perceived, thus minimizing the time between configuration and analysis of a simulation. With advances in dense sensing modalities and experiential media, previous responsive media systems have expanded upon these primarily screen-based interactions in several aspects:

1. Embodied, enactive environments allow comparatively unconstrained engagement with the computation, such as through full-body movement or the use of physical props or other aspects of the environment. For example, gestural input can afford more degrees of freedom to allow multiple parameters to be controlled simultaneously and physical props can be used to construct detailed geometry for more varied, non-parametric boundary conditions.
2. Applications range widely from basic experiential experiments (e.g. relation between memory and corporeal movement, or rhythmic entrainment of ensembles of people and time-based media processes) to artistic installations and performance (e.g. Serra [61] and Timelenses [62]).[2]
3. Designing for whole body and ensemble engagement implies *thick* (Sha [5]: 72), multimodal, analog and digital engagement with the environment as opposed interacting along one or a few dimensions of sensory perception.

In studying the use of responsive media environments for computational scientific investigation, our key interest lies in observing what types of behaviors in these systems could contribute to a scientific practice, such as rapidly "sketching" hypotheses, perhaps in advance of more numerically reproducible studies. With lowering costs and advances in computing resources, both in HPC systems and desktop hardware, many of the simulations that now require HPC systems may eventually be able to be steered at responsive rates, so we study the use of those models which can currently be simulated at these rates to get an understanding of where responsive interaction can fit into future computational science workflows.

[2] See videos https://vimeo.com/synthesiscenter/serra2017 and https://vimeo.com/synthesiscenter/palimpsest.

2 EMA: An Experiential Model of the Atmosphere

As an initial exploration, EMA is installed in the Intelligent Stage ("iStage") space in Synthesis at the School of Arts, Media, and Engineering at Arizona State University: a 30 × 30-foot black box space with a sprung dance floor, theater grid with 16 DMX-controllable RGB LED theatrical lights and additional floor-mounted lights, 4K floor and vertical scrim LED projections, horizontal ceiling-mounted and vertical infrared-filtered cameras, infrared light emitters, floor-mounted contact and boundary microphones, grid-mounted microphones, and 8.2-channel surround audio. The space is designed to be modular to support multiple responsive environments with flexible HD video routing and AVB audio routing hardware. The Synthesis research team has created a suite of software tools in Max/MSP/Jitter, known as SC^3, for animating the space for creating responsive environments, allowing both novice and expert developers to create new environments using a suite of software frameworks and abstractions. This is refined from multiple generations of researchers working on predecessor "media choreography" composition systems for responsive environments (Sha [5,63]).

As a responsive, steerable model of warm cloud physics, EMA satisfies many of the objectives of our research into dense computational media that can be steered through collaborative human gesture and physical configuration of the space. Additionally, using a physical model of atmospheric dynamics allows us to explore human interaction with a simulated physical model that leverages participants' existing physical intuition of matter, exhibits phase changes, and can simulate phenomena at different spatial and temporal scales, all contributing to a rich set of processes and forms that can be studied by investigators in the space.

EMA implements an incompressible fluid flow model along with additional computation of buoyancy, condensation and evaporation of water vapor, and thermodynamics. During each timestep of the simulation, the model allows for external video textures to manipulate the simulated fields, including air velocity, pressure, water vapor, liquid water, temperature, and viscosity. Global scalar parameters can also be manipulated in real-time, such as ground pressure and temperature, altitude, spatial and temporal scale, specific heat capacities of dry air and water vapor, external wind speed and direction, and gravity magnitude and direction. For mathematical and implementation details, see [64].

2.1 Visualization

The model's fields can then be viewed with a number of different visualization modes, including conventional pseudocolor images given a colormap, particle flow fields, tracer particles, line integral convolution, vector feather plots, and additional artistic renderings composed of multiple fields, such as temperature and different phases of water. The base set of visual mappings has been developed to reflect conventional scientific visualizations from familiar platforms such as

[3] See http://vimeo.com/synthesiscenter/demo.

Matplotlib, ParaView, and MATLAB. Each of these visualization modes can be seemlessly interchanged using a mobile tablet interface, and EMA supports layering multiple visualizations, such as being able to view flow lines or specific tracer particles on top of a composite rendering of air temperature, water vapor, and condensed liquid water. The tablet interface also allows viewers to adjust scaling parameters, color maps, and compositing in situ without the need to return to a desktop interface in order to encourage all investigative activity to occur embedded within the simulation environment.

2.2 Sonification

The sonic affordances of the space can also be used to communicate important dynamics within the simulation that may be difficult to attend to visually, such as spatialized activity of air flow or the position and velocity of moving tracer particles. We have implemented two modes of sonification in the environment that allow investigators to sonify activity within regions and particular points in space. In particular, a field-based sonification tool allows multiple participants to scale a bounding rectangle around their bodies or static objects to sonify the dynamics of specific regions of space. The underlying field is then subdivided into a variable number of zones, and the average changes of the field within the zones are then sent through a multi-channel sample player and filterbank and ambisonically spatialized around the participants [64].

In a separate particle-based sound synthesizer, individual tracer particles are simulated in the model, which follow the velocity of the air. Each particle is mapped to a separate voice, and its speed and direction are mapped onto different aspects of the synthesized sound. To allow designers or investigators to choose informative sound textures, they are able to select an audio file or recorded audio sample which is then sampled with a granular synthesizer. The angle of velocity of each tracer particle is then mapped to the center frequency of a resonant bandpass filter, and the speed of the particle is mapped to the particle's volume. This mapping is particularly effective at sonifying sudden changes in particle velocity, such as when it enters a vortex or suddenly encounters a gust of wind. When a particle is in circular motion, for example, the synthesized voice will make repeating sweeps up and down in frequency content. Each particle is then ambisonically spatialized within the space to allow participants to understand whole-field dynamics when they are visually focused on a particular region of the simulation.

3 Enactive Scenarios

The basic dimensions of our enactive scenarios are the number of people and props (physical manipulables that can be tracked), the weather phenomenon being simulated and experienced, the tools or instruments for inspecting or modulating the state of the simulation.

As earlier mentioned, we distinguish between the experience of one, two, or ensembles ($n \geq 3$) of people co-constructing an experience in realtime with the steerable environment. Thus the experiments on how the simulation is experienced are designed differently and accordingly. For each of three scenarios, we list recorded experimental behavior from open-ended sessions working with the model as a solo investigator, a pair, a pair using objects in the lab space to construct experiments, and as a guided ensemble. In the ensemble scenario, people dispensed with instruments and used their bodies to walk in a coordinated way to steer the simulation holistically. These scenarios include:

– **Cloud formation and air flow on a horizontal plane**: a horizontal simulation with a ceiling-mounted camera is constructed where each square pixel corresponds to 900 m^2 of simulated space, the simulated ambient temperature is 150 Kelvin, and motion of entities in the space is mapped to an increase in water vapor, which condenses nearly instantaneously.
– **Cloud formation on a vertical plane**: a simulation with a vertically oriented camera facing an opposing vertical projection surface is constructed where each pixel corresponds to 100 m^2 of simulated space and the lapse rate of ambient temperature with altitude is 6.5 K/km with a sea level temperature of 288.15 K, resulting in a temperature gradient ranging from 288.15–241.35 K. Presence of bodies and objects in the space acts as an obstruction to fluid flow, while movement is mapped to an increase in water vapor and temperature, causing buoyant lift and eventual condensation, usually slightly above-head when participants are standing approximately 5 ft from the projection.
– **Cloud formation and air flow with wind on a vertical plane**: a simulation parameterized similarly to the previous scenario, but an external, constant source of downstage velocity (wind) is added, allowing participants to observe the effects of air flow around themselves and objects.

Table 1 summarizes observations of novel, investigative participant behavior with the simulation in three different scenarios. Increasing the number of participants in the space can be seen to increase joint expressive capabilities, such as through actions as coordinated movement and manipulation of instruments and sharing objects or physical space in the eye of the camera. Inclusion of physical instruments in the space, ranging from isolated objects such as pipes and rope to furniture, such as stools and tables, allows participants to construct stable fluid boundaries or affect larger and more complex regions of the simulation than they could with their bodies alone, such as through spinning objects overhead or jointly moving large objects between each other.

4 Participant Response

When there are two or more people in the space, we can use "second person" elicitation techniques where instead of providing the participants with a predesigned set of descriptors and metrics from which to choose a classification of their experience, we ask them to converse with one another and come up

Table 1. Observed experimentation strategies in three simulations.

<table>
<tr><th>Fluid flow and cloud formation on a horizontal plane</th><th>Cloud formation on a vertical plane</th><th>Cloud formation and air flow with wind on a vertical plane</th></tr>
<tr><td colspan="3">One person</td></tr>
<tr><td>Spinning with arms outstretched to produce vortices, clouds</td><td>Swaying side-to-side to introduce buoyant water vapor and heat and watch clouds form</td><td rowspan="3">Raising one or both hands to obstruct wind above lifted condensation level</td></tr>
<tr><td>Using sheets of paper, mapped to low pressure centers, to direct air flow</td><td>Raising one or both hands and moving them parallel to the screen to introduce water vapor above the lifted condensation level</td></tr>
<tr><td>Walking in circles to produce vortices</td><td>Walking parallel to the projection to leave a trail of buoyant water vapor</td></tr>
<tr><td colspan="3">Two people</td></tr>
<tr><td>Walking together in circles to produce more stable vortices</td><td rowspan="2">Holding or overlapping hands and moving hands down to leave a large trail of buoyant water vapor</td><td>Standing in order of shortest to tallest or vis versa to observe flow up or down an irregular slope</td></tr>
<tr><td>Walking towards each other with arms outstretched to collide opposing fronts</td><td>Using side-by-side hands at different heights to create irregular slopes</td></tr>
<tr><td colspan="3">Two people with instruments</td></tr>
<tr><td>Manipulating pipes to create an obstruction with a variable-sized slit or channel to observe effects on vortex sheets</td><td rowspan="5">Rotating and moving a vertically oriented foam tube horizontally to produce large areas of buoyant water vapor</td><td>Rotating a large foam rectangle to observe flow up a straight incline</td></tr>
<tr><td>Shaping an airfoil with a rope and observing effects on fluid pressure and velocity on either side</td><td rowspan="4">Placing a sheet of dark paper close to the camera to observe flow up a straight incline</td></tr>
<tr><td>Creating different curves and shapes with a rope at an oblique angle to simulated wind and observing fluid flow along their surface</td></tr>
<tr><td>Moving circular tables and stools into the space to observe flow around many objects</td></tr>
<tr><td>Spinning a wand overhead to create vortices</td></tr>
<tr><td colspan="3">Ensemble $n \geq 3$</td></tr>
<tr><td>Walking together in circle to produce a stable vortex</td><td rowspan="2">Producing cloud masses as concurrent effect</td><td rowspan="2">Producing cloud masses as concurrent effect</td></tr>
<tr><td>Participating in group discussion and manipulation of several objects (sheets of black paper, rope, metal tubes) to experiment and discuss results of different configurations on fluid flow</td></tr>
</table>

with a commonly agreed upon account of what they experienced, and what they thought was happening. This phenomenologically informed experimental method has been elaborately developed by C. Petitmengin and colleagues for getting shared, and thus socially objective accounts of thick experience (Petitmengin [27, 65, 66]).

As an example of coordinated activity with shared physical instruments, in Dialog 1, two investigators, given names A and B, work within the fluid flow and cloud formation on a horizontal plane scenario, visualizing the flow velocity with a mode mapping angle to hue and magnitude to intensity. An external source of wind flows downstage. Within the course of 20 min, the two participants constructed several different shapes to act as obstacles to fluid flow and propose and test hypotheses regarding the relationship of the size and periodicity of vortex sheets to the size of a gap between obstacles. Being able to pause the simulation allowed for more thorough examination of a phenomenon, and using multiple visualization types allowed the participants to test a hypothesis about the relationship of fluid velocity and pressure surrounding an object. Additionally, the experience prompted further discussion about a specific topic that would continue outside the scenario.

In ensemble work, with three to on the order of 40 people, the facilitator guided joint activities by suggesting coordinated or disaggregated movement. We are learning to exploit the unique features of having a large common space in which large numbers of participants can jointly steer the simulation without props simply by coordinating their whole-body interaction with EMA. One persuasive and enlightening instance of such coordinated steering is when participants walked in a ring to create and move a common vortex (hurricane) while the vertical projection showed warm air condensing into clouds.

5 Conclusions

From our early trials using the Experiential Model of the Atmosphere, we have demonstrated the potential of responsive environments, which can respond equally to the activity of individuals, ensembles of people, and physical objects and other entities within the space, to find use in computational science practice. Moving beyond traditional single-user WIMP interfaces and into embodied, enactive computing environments opens up many new possible interaction modalities, as simulations can be interactively steered using gestural interaction, where people themselves are the scientific instruments, and through quickly prototyping and manipulating physical instruments as extended interfaces.

From our recorded experimentation sessions, we have witnessed novel gestural interaction, improvised coordination amongst individuals and within groups, on-the-fly construction of instrumentation, and abductive hypothesis formation and testing. Providing an enactive environment that is conducive to comparatively unconstrained exploration of a model compared to traditional interfaces or simulation parameterization scripts, that is *play* (Huizinga [67], Sutton-Smith [68], Sha [5]), can be a productive step in eliciting original scientific thought in computational scientific practice.

A: I was thinking we could use these pipes to see flow through a small slit.

A hands B a 4-foot metal pipe and picks up a similar pipe.

B: OK.
A: So maybe we can put these going this way and leave a gap between them.

Both place the pipes along a line bisecting the stage, orthogonal to the simulated wind and with a small gap between them.

A: So it seems to me the air is flowing right through the slit without being affected much, right?
B: Yeah, it's the same shade of red.
A: I feel like we should see a vortex sheet. I wonder if the visualization's just scaled so that we can't see the vertical movement. Maybe it's too small.
B: Ah, maybe.
A: Maybe I can try lowering the vertical scale of the velocity so we can see it more.
B: OK.

A uses the tablet interface to adjust the visualization scaling and then changes it back again.

A: Oh wait, I know. Do you think moving [the pipes] closer together could help?
B: Yes, maybe. Let's try it.

Each person moves to either end of the pipes and pushes them closer together, watching the simulation on the floor to gauge their distance.

B: Ah! Now we see them. Maybe before the first oscillation would have been off the stage.
A: Oh! Maybe. So maybe the oscillations spread out with a larger gap?
B: Let's try slowly pulling them apart.

They move the pipes slightly apart.

A: It seems like there are fewer oscillations now. Hmm, I guess we could test this by pausing the simulation and measuring with a tape measurer.

[Later ...]

B: What if we rotate [the pipes]?
A: How?
B: Make a 'V.' So put yours on that end, leaning toward me.

Both place the pipes at an angle, with a slight gap downstage, in the direction of the fluid flow.

A: Ah ha. Seems like it's behaving pretty similarly. The vortex sheet comes out of the tip.

B: Wait, I actually have a book that talks about these.

B goes to their computer.

[Later ...]

B: Yeah. Maybe we can try different shapes with that rope you have.
A: I was thinking we could test out an airfoil. Like this . . .

A shapes an airfoil out of the rope.

A: You can sort of see lower velocity on the top compared to the bottom, which makes sense . . . it should have higher pressure too, I think.
B: How do you see the pressure?

A uses the tablet to switch to a pressure visualization.

B: So the lighter color is higher pressure?
A: Yeah. Seems to be working. There's higher pressure against all the edges that are getting hit, but there's a larger region of high pressure coming off the leeward side.

Dialog 1: Two participants experiment in the horizontal simulation scenario, using pipes and a rope to construct different boundaries for fluid flow.

Acknowledgments. We thank the faculty, technical staff, and student researchers with whom we have worked so closely since the establishment of the Synthesis Center: Todd Ingalls, Peter Weisman, Connor Rawls, Garrett L. Johnson, Assegid Kidane, Megan Patzem, Keyaanna Pausch, and Emiddio Vasquez. Additionally, we thank our consulting scientific collaborators Melissa Bukovsky, Randy Russell, and Linda Mearns, whose early generous involvement has shaped much of our thinking of integrating these tools into scientific computing workflows.

References

1. Varela, F.J., Thompson, E., Rosch, E.: The Embodied Mind: Cognitive Science and Human Experience. MIT Press, Cambridge (1991)
2. Dourish, P.: Where the Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2001)
3. Sha, X.W.: Calligraphic video: using the body's intuition of matter. Int. J. Creat. Interfaces Comput. Graph. **1**(1, Inaugural Issue) (2009)
4. Sha, X.W., Freed, A., Navab, N.: Sound design as human matter interaction. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2013, pp. 2009–2018. ACM, New York (2013)
5. Sha, X.W.: Poiesis and Enchantment in Topological Matter. MIT Press, Cambridge (2013)
6. Ingalls, T.: Affect in media: embodied media interaction in performance and public art. IEEE Multimedia **20**(3), 4–7 (2013)

7. Merleau-Ponty, M., Landes, D.A.: Phenomenology of Perception. Routledge, Abingdon (2011)
8. Casey, E.S.: Borders and Boundaries: Edging into the Environment, Merleau-Ponty and Environmental Philosophy: Dwelling on the Landscapes of Thought. SUNY Press, Albany (2007). Ed. by S.L. Cataldi and W.S. Hamrick
9. Casey, E.S.: Getting Back Into Place: Toward a Renewed Understanding of the Place-world. Studies in Continental Thought, 2nd edn. Indiana University Press, Bloomington (2009)
10. Morris, D.: The Sense of Space. SUNY Series in Contemporary Continental Philosophy. State University of New York Press, Albany (2004). (2004045291 David Morris. ill.; 24 cm. Includes bibliographical references (pp. 197–213) and index)
11. James, J., Ingalls, T., Qian, G., Oldon, L., Rikakis, T.: Lucidity: a movement - based interactive dance performance. E-journal LEA **15**(5), 1–8 (2007)
12. James, J., Ingalls, T., Qian, G., Olsen, L., Whiteley, D., Wong, S., Rikakis, T.: Movement-based interactive dance performance. In: Proceedings of 14th Annual ACM International Conference on Multimedia - MULTIMEDIA 2006, p. 470. ACM Press, New York (2006)
13. Ingalls, T., James, J., Qian, G., Thornburg, H., Campana, E., Rajko, S., Mumford, J.: Response: an applied example of computational somatics (2007)
14. Wallis, I., Ingalls, T., Campana, E., Vuong, C.: The instrumental hobby and long-term engagement with HCI. In: 25th BCS Conference on HumanComputer Interaction. Number 1975, British Computer Society (2011)
15. Baran, M., Lehrer, N., Duff, M., Venkataraman, V., Turaga, P., Ingalls, T., Rymer, W.Z., Wolf, S.L., Rikakis, T.: Interdisciplinary concepts for design and implementation of mixed reality interactive neurorehabilitation systems for stroke. Phys. Ther. **95**, 449–460 (2014)
16. Chen, Y., Huang, H., Xu, W., Wallis, R.I., Sundaram, H., Rikakis, T., Ingalls, T., Olson, L., He, J.: The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. In: Proceedings of 14th Annual ACM International Conference on Multimedia, pp. 763–772. ACM (2006)
17. Mearns, L., Arritt, R., Biner, S., Bukovsky, M., Cay, D., Flory, D., Gutowski, W., Jones, R., Moufouma-Oki, W., Laprise, R., Leung, R., Qian, Y., McGinnis, S., McDaniel, L., Middleton, D., Sain, S., Nunes, A., Roads, J., Sloan, L., Snyder, M., Takle, G.: North American regional climate change assessment program: overview of phase 1 results. Bull. Amer. Metr. Soc. **93**, 1337–1362 (2011)
18. Mearns, L.O., Bukovsky, M.S., Pryor, S.C., Magaña, V.: Downscaling of climate information. In: Ohring, G. (ed.) Climate Change in North America. RCS, pp. 201–250. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-03768-4_5
19. Bukovsky, M., Gochis, D., Mearns, L.: Towards establishing NARCCAP regional climate model credibility for the North American monsoon: current climate simulations. J. Climate **26**, 8802–8826 (2013)
20. Warner, T.T.: Numerical Weather and Climate Prediction. Cambridge University Press, Cambridge (2011)
21. Magnussen, T.: Of epistemic tools: musical instruments as cognitive extensions. Org. Sound **14**, 168–176 (2014)
22. James, W.: Essays on Radical Empiricism. Dover (Harvard 1912) (2003 (1912))
23. Gendlin, E.T.: Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective. Northwestern University Press, Evanston (1997) (Eugene T. Gendlin. 23 cm. Northwestern University studies in phenomenology and existential philosophy Originally published: Glencoe: Free Press of Glencoe, c1962. With new pref)

24. Casey, E.S.: The Fate of Place: A Philosophical History. University of California Press, Berkeley (1997)
25. Casey, E.S.: Perceiving and remembering. Rev. Metaphys. **32**, 407–436 (1979)
26. Casey, E.S.: Remembering: A Phenomenological Study, 2nd edn. Indiana University Press, Bloomington (2000)
27. Petitmengin, C.: Researching the dynamics of intuitive experience. In: Handbook of Research Methods on Intuition, pp. 188–198. Edward Elgar Publishing, Boston (2014). (Ed. by, M. Sinclair)
28. Sheets-Johnstone, M.: From movement to dance. Phenomenol. Cogn. Sci. **11**(1), 39–57 (2012)
29. Sheets-Johnstone, M.: The Primacy of Movement. John Benjamins Publishing Company, Amsterdam (1999)
30. Sheets-Johnstone, M.: Thinking in movement. J. Aesthet. Art Crit. **39**(4), 399–407 (1981)
31. Petitmengin, C.: Le corps du savant à la source du sens. Epistémologie du corps du savant. pp. 188–198. L'Harmattan, Paris (2014). (Ed. by, M. Quidu)
32. Sha, X.W., Gill, S.: Gesture and response in field-based performance. In: 5th conference on Creativity and Cognition, Ernest Edmonds, University of Technology, Sydney, vol. 5, pp. 205–209. ACM
33. Sha, X.W.: Resistance is fertile: gesture and agency in the field of responsive media. Configurations **10**(3), 439–472 (2002)
34. Bregman, A.: ASA at McGill, Keynote CIRMMT, CIRMMT. McGill University, Montreal (2008)
35. Sundaram, H.: Experiential media systems. ACM Trans. Multimedia Comput. Commun. Appl. **2**(3) (2013)
36. Rikakis, T., He, J., Sundaram, H., Qian, G., Spanias, A.: An interdisciplinary arts and engineering initiative for experiential multimedia. In: ASEE Annual Conference and Exposition, ASEE, pp. 8715–8724 (2005)
37. Wei, S.X.: Poetics of performative space. In: Gill, S. (ed.) Cognition Communication and Interaction, pp. 549–566. Springer, London (2008). https://doi.org/10.1007/978-1-84628-927-9_31
38. Bullivant, L.: Responsive Environments: Architecture, Art and Design. Paperback edn. Victoria and Albert Museum (2006)
39. Massumi, B.: Parables for the Virtual: Movement, Affect, Sensation (Post-Contemporary Interventions). Duke University Press (2002)
40. Longo, G., Montévil, M.: Perspectives on Organisms: Biological Time, Symmetries and Singularities. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-35938-5
41. Simondon, G.: On the mode of existence of technical objects. Univocal. Minnesota (2016 (1958)). Translated by C. Malaspina and J. Rogove
42. Scott, D.: Gilbert Simondon's Psychic and collective individuation: a critical introduction and guide. Edinburgh University Press (2014)
43. Combes, M.: Gilbert Simondon and the Philosophy of the Transindividual. Technologies of Lived Abstraction. MIT Press, Cambridge (2013). 2012013224 Simondon. English Muriel Combes; translated, with preface and afterword, by Thomas LaMarre. 24 cm. Includes bibliographical references (pp. 109-119)
44. Maturana, H.R.: Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. Cybern. Hum. Know. **9**(3–4), 5–34 (2002)

45. Maturana, H.R., Varela, F.J.: Autopoiesis and Cognition: The Realization of the Living. Boston Studies in the Philosophy of Science; v. 42. D. Reidel Pub. Co./Holland, Dordrecht/Boston (1980). H.R. Maturana and F.J. Varela; with a pref. to "Autopoiesis" by Sir Stafford Beer. 23 cm. Includes index
46. Jameson, F.: Brecht and Method. Verso, London (2000)
47. Baran, M., Lehrer, N., Siwiak, D., Chen, Y., Duff, M., Ingalls, T., Rikakis, T.: Design of a home-based adaptive mixed reality rehabilitation system for stroke survivors. In: Proceedings of 33rd Annual International Conference of IEEE Engineering in Medicine and Biology Society, vol. 2011, pp. 7602–7605, January 2011
48. Morris, D.: Bergsonian intuition, husserlian variation, peirceian abduction: toward a relation between method, sense, and nature. South. J. Philos. **43**, 267–298 (2005)
49. Peirce, C.S.: Pragmatism as the Logic of Abduction (Lecture VII of the 1903 Harvard Lectures on Pragmatism). Volume Essential Peirce v. 2, pp. 226–241
50. Psillos, S.: Abduction: between conceptual richness and computational complexity. In: Flach, P.A., Kakas, A.C. (eds.) Abduction and Induction: Essays on their Relation and Integration. APLS, vol. 18, pp. 59–74. Springer, Dordrecht (2000). https://doi.org/10.1007/978-94-017-0606-3_4
51. García, M., Duque, J., Boulanger, P., Figueroa, P.: Computational steering of CFD simulations using a grid computing environment. Int. J. Interact. Des. Manuf. (IJIDeM) **9**(3), 235–245 (2015)
52. Slotnick, J.P., Khodadoust, A., Alonso, J.J., Darmofal, D.L., Gropp, W.D., Lurie, E.A., Mavriplis, D.J., Venkatakrishnan, V.: Enabling the environmentally clean air transportation of the future: a vision of computational fluid dynamics in 2030. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **372**(2022), 20130317 (2014)
53. Konev, A., Waser, J., Sadransky, B., Cornel, D., Perdigão, R.A., Horváth, Z., Gröller, M.E.: Run watchers: automatic simulation-based decision support in flood management. IEEE Trans. Vis. Comput. Graph. **20**(12), 1873–1882 (2014)
54. Cornel, D., Konev, A., Sadransky, B., Horváth, Z., Gröller, E., Waser, J.: Visualization of object-centered vulnerability to possible flood hazards. Comput. Graph. Forum **34**(3), 331–340 (2015)
55. Sheharyar, A.: Parallelization of computational steering of the electron avalanches in the high-energy particle detector simulators. Masters thesis, Qatar University (2014)
56. Xiao, J., Zhang, J., Yuan, Y., Zhou, X., Ji, L., Sun, J.: Hierarchical visual analysis and steering framework for astrophysical simulations. Trans. Tianjin Univ. **21**(6), 507–514 (2015)
57. Zhang, S., Xie, S., Wang, K., Yuan, Y.: Multi-scale cardiac electrophysiological simulation: a GPU-based system by computational steering. ISET 160–163 (2015)
58. Parker, S.G., Johnson, C.R.: SCIRun: a scientific programming environment for computational steering. In: Proceedings of 1995 ACM/IEEE Conference on Supercomputing, p. 52. ACM (1995)
59. Brooke, J.M., Coveney, P.V., Harting, J., Jha, S., Pickles, S.M., Pinning, R.L., Porter, A.R.: Computational steering in realitygrid. In: Proceedings of UK e-Science All Hands Meeting, vol. 16 (2003)
60. Nguyen, H.A., Abramson, D., Kipouros, T., Janke, A., Galloway, G.: Workways: interacting with scientific workflows. Concurr. Comput.: Pract. Exp. **27**(16), 4377–4397 (2015)
61. Khintirian, O.S.: Serra vegetal life (2016). http://serracreation.weebly.com/
62. Stein, J., Khintirian, O.S., Ingalls, T., Sha, X.W.: Time lenses apparatus (2016). http://rhythmanalysis.weebly.com/time-lenses-apparatus.html

63. Sha, X.W., Fortin, M., Navab, N., Sutton, T.: Ozone: a continuous state-based approach to media choreography (2010)
64. Mechtley, B., Stein, J., Roberts, C., Wei, S.W: Rich state transitions in a media choreography framework using an idealized model of cloud dynamics. In: Thematic Workshops 2017 - Proceedings of Thematic Workshops of ACM Multimedia 2017, Co-located with MM 2017 (2017)
65. Petitmengin, C., Bitbol, M.: The validity of first-person descriptions as authenticity and coherence. In: Petitmengin, C. (ed.) Ten Years of Viewing from Within: The Legacy of Francisco Varela. Imprint Academic, Exeter (2009)
66. Petitmengin, C.: Towards the source of thoughts, the gestural and transmodal dimension of lived experience. J. Conscious. Stud. **14**(3), 54–82 (2007)
67. Huizinga, J.: Homo Ludens: A Study of the Play-Element in Culture. Beacon Press, Boston (1971)
68. Sutton-Smith, B.: The Ambiguity of Play. Harvard University Press, Cambridge (2001). A study of what it means to play. The book explores play theory as elaborated and debated in disciplines from biology and psychology to mathematics, and analyzes the implications of play in areas such as child development and the Western work ethic, examining the values dictating forms of play

Reconstruction by Low Cost Software Based on Photogrammetry as a Reverse Engineering Process

Dolores Parras(✉), Francisco Cavas-Martínez, José Nieto, Francisco J. F. Cañavate, and Daniel García Fernández-Pacheco

Graphical Expression Department, Technical University of Cartagena, 30202 Cartagena (Murcia), Spain
{dolores.parras,francisco.cavas,jose.nieto, francisco.canavate,daniel.garcia}@upct.es

Abstract. Among the various types of scanning that can be found on the market to perform a three-dimensional reconstruction, an alternative is highlighted due to its low cost and its ease of use, making it suitable for a great amount of applications. This is Image-based 3D Modeling and Rendering (IB3DMR), in which it is possible to generate a three-dimensional model from a set of 2D photographs. Among the existing commercial applications based on the IB3DMR, this communication has selected the Autodesk ReCap software, which is free and provides great features in terms of simplicity of operation, automation of the reconstruction process and possibility of exporting to other more complex applications. The use of this type of technologies based on photogrammetry is an alternative to the conventional reverse engineering processes, so a study with seven different pieces in terms of colour, geometry and texture has been performed for its assessment, obtaining three-dimensional reconstructions with very satisfactory results.

Keywords: 3D reconstruction · Virtual model · Image processing

1 Introduction

The reverse engineering process is based on the obtaining of information from a product to know its operation, the material with which it was made or to obtain a digital model. Unlike the general design process, in which technical data is used to produce a particular product, in reverse engineering the process is the opposite. Reverse engineering relies mainly on a data acquisition system, where computer vision has become a vital part for extracting useful information from reality. These advances have allowed computers to perceive reality through sensors based on optical principles. These systems, commonly called scanners, can be two-dimensional, providing flat images, or three-dimensional, digitalizing volumes.

A great variety of three-dimensional scanners providing a great data accuracy can be found on the market: through contact, without contact, active, passive, etc. [1, 2].

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 145–154, 2018.
https://doi.org/10.1007/978-3-319-91581-4_11

This type of technology is not always easy to acquire or to use due to its high cost and complexity. As a result, more and more affordable software-based alternatives are being used. These new alternatives consist of three-dimensional scanning and reconstruction of an object from a series of photographs, which is known as Image-Based 3D Modeling and Rendering (IB3DMR). More specifically, this method is based on the detection, grouping and extraction of image features (sides, vertices, etc.) that are later interpreted as common elements of a 3D model. The IB3DMR method covers from data acquisition to the generation of an interactive 3D virtual model, a process in which a point cloud is converted into a triangulated network (mesh) or a textured surface.

There are several software alternatives on the market that provide 3D reconstruction using images, such as Autodesk ReCap (new version of 123D Catch and Remake), Microsoft PhotoSynth, Patch-based Multi-View Stereo (PMVS), PixelStruct, VisualSize, Bundler, BigSFM or OpenPhotoVR, among others. Previous studies comparing the reconstruction results obtained by some of these applications have been carried out. Cavas Martínez et al. [3] studied the behaviour of Microsoft PhotoSynth and PMVS for the reconstruction of different types of fruits, offering the PMVS software better results when applying textures. A subsequent study by the same authors [4] analysed Microsoft PhotoSynth, PMVS and 123D Catch software for the reconstruction of metallic mechanical parts, demonstrating the high reconstruction precision obtained by 123D Catch versus the rest; PMVS does not give good results when textures are applied, and Microsoft PhotoSynth only provides a few point cloud that does not completely conform the scanned piece.

One of the first works that demonstrated the performance of Autodesk 123D Catch was realized by Cervera et al. [5], who chose 123D Catch along with other applications to perform the reconstruction of an acoustic study, performing indoor 3D modelling tests. In the field of architecture several works that study possible measurement errors between different architectural elements and their image-based 3D reconstructions can be found [6], as well as other works in which reconstructions obtained by laser methodology and by using images are compared [2, 7] or in which restoration elements are reconstructed [8]. Fields of application for this technology can be very diverse, finding studies in archaeology [9], in odontology for 3D reconstruction of human faces or complementing others diagnostic tests [10], for 3D modelling of terrains [11, 12], for reconstructing cave paintings [13], or even under the sea, where this methodology has been used to obtain the reconstruction of aquatic organisms in an easy, fast and non-intrusive way [14].

On the basis of the results provided by these studies, the Autodesk ReCap software tool, which is considered the new version of Autodesk 123D Catch and Autodesk Remake [15], has been selected for this work. The aim of this study is to validate the Autodesk ReCap application as a low cost tool for generating 3D reconstructions as part of the reverse engineering process. For this objective, the digitization of several pieces with different features in terms of geometry, materials and textures has been carried out, assessing their suitability to this process.

2 Materials and Methods

2.1 The Reverse Engineering Process

Two paths to obtain virtually objects and their subsequent use for possible redesign or prototype fabrication can be defined within the reverse engineering process. The traditional scheme for data acquisition of an existing object and its possible reproduction by means of prototype printing is shown in Fig. 1.

Fig. 1. Data acquisition scheme using measuring tools.

However, an alternative to this process can also be used: the scanning of objects by using images and their transformation into virtual models with which to work and obtain the prototypes (Fig. 2).

Fig. 2. Data acquisition scheme using photographs.

Thus, a reverse engineering process with the aid of a low-cost computer application for the digitization of everyday pieces is defined below, distinguishing the different stages followed to obtain the three-dimensional reconstructions: (i) scanning, (ii) reconstruction, and (iii) 3D printing.

2.2 Phase I: Scanning

Image capture was performed in a dynamic environment, in which the pieces remained fixed and the photographs were taken around them and at two different heights (0° and 30° approximately). Between 30 and 50 photographs were taken for each piece.

A Nikon D80 Digital SLR Camera with a 10 megapixel resolution and a calibrated sensitivity of 100 to 1600 ISO was used for all images captured during the study. For image processing, a computer endowed with an Intel® Core ™ i7-4790k 4 GHz processor, 16 GB of RAM memory and Windows 7 Professional 64-bit operating system was used.

With the aim of evaluating the suitability of using this type of technology for the digitalization of pieces, seven pieces with different shape and texture were selected to perform the scanning and reconstruction phases (see Fig. 3): a dragon's head, a helmet, a toy car, a computer mouse, an electric screwdriver, a cube of colours and a miniature Eiffel Tower.

Fig. 3. Seven pieces used during the study.

2.3 Phase II: Reconstruction

For the three-dimensional reconstruction of the objects, Autodesk ReCap software was used. The reconstruction process begins importing the photographs taken in the previous scanning phase. The photographs are uploaded to the Autodesk servers, where the reconstruction is created and can be later downloaded and displayed in the application. In this way, virtual models present a good quality and are processed quickly, without consuming CPU or RAM resources from the user computer and depending only on the Internet connection speed.

Once the process is finished, the final result is displayed on the screen (Fig. 4). After verifying the quality of the reconstructed object, a cleaning process for those environment parts surrounding the object must be executed. Then, the model can be exported to any of the five formats offered by the application: OBJ, FBX, STL, PLY and PTS. For the present study, the generated models were exported to STL format.

As an example, Fig. 4 shows the reconstructions of two objects provided by the application once the model has been downloaded (left), as well as the edited model where the entire environment has been removed (right).

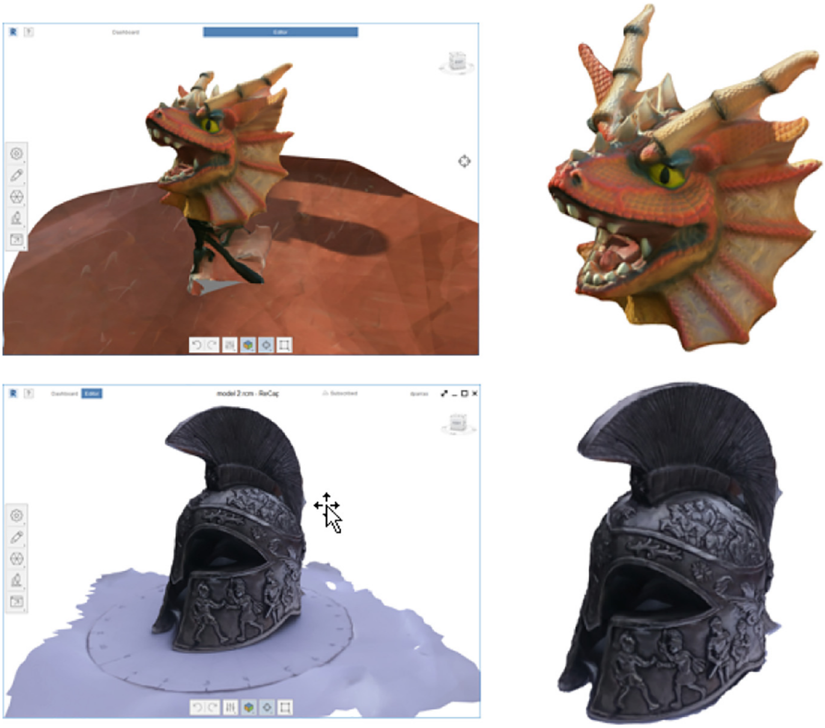


Fig. 4. 3D reconstruction with background (left) and without background (right).

2.4 Phase III: 3D Printing

One of the reasons for obtaining virtual models is their subsequent reproduction using 3D printing. This process consists of making almost any shape three-dimensional solid object from digital models in STL format. The STL format permits the geometry of 3D objects to be defined, excluding information such as colours, textures or physical properties included in other CAD formats. For this reason, it is the standard format for additive manufacturing technologies. It uses a closed triangular mesh to define the shape of an object. The smaller the triangulation size is used, the higher final resolution will be obtained. In this sense, Autodesk ReCap offers the possibility of exporting the virtual model into STL format to be sent where it will be later printed. Figure 5 shows the mesh surfaces of the model in Autodesk ReCap with texture (left) and without texture (middle), as well as the model in STL format in the MeshLab software, which is a low cost application for analysis and edition of meshes. This tool is used to optimize, if necessary, the virtual model mesh, giving the possibility of exporting to other formats to be used in CAD applications.

Fig. 5. Mesh surfaces with texture (left) and without texture (middle) in Autodesk ReCap, and model in STL format in Meshlab (right).

3 Results

The results of the reconstructions obtained during the study were very satisfactory in most cases, being considered very suitable for their use in any required digital format. Moreover, prototypes by 3D printing can be obtained from the virtual models generated, if necessary. It must be taken into account that slimmer figures, such as the miniature Eiffel Tower example, have a less defined finishing, which may require paying more attention in the scanning phase. Lighting is an important feature to be considered during the scanning process, since a homogenous illumination during the image capture greatly favours the reconstruction quality.

Table 1 shows a comparison between the real images of each piece and the reconstructions obtained using this methodology. An additional video file showing the reconstructions of all pieces is also provided as complementary material for this publication.

Table 1. Comparative table with reconstruction results for all studied pieces

Model	*Real piece*	*3D reconstruction*
Dragon´s head		
Helmet		
Toy car		
Computer mouse		
Electric screwdriver		
Cube of colours		
Miniature Eiffel tower		

4 Discussion

Digitization of parts is becoming a challenge in a wide variety of fields of application. The main problem with these reverse engineering processes is the lack of knowledge of existing technologies. Most of them are sophisticated, not accessible, high-priced and require some previous training. For this reason, there is an economical, flexible and easy to use alternative that permits getting started in this type of data acquisition and reconstruction processes: software based on photogrammetry. Available software based on photogrammetry allow performing scans and 3D reconstructions with a very good quality and a very low cost.

Software based on photogrammetry has experienced a great evolution since its inception, optimizing the input data and obtaining higher quality results of reconstruction. It should be pointed out that this quality strongly depends on the scanning phase, where a series of 2D photographs are taken around the object considering several parameters such as lighting, focus, good distribution of photographing, etc.

After analysing different low-cost applications based on photogrammetry, Autodesk ReCap was selected for this study and the recommendations provided by the software developer for taking photographs of all objects were followed, that is to say, letting the piece fixed and moving the camera around it covering 360°. The photographs were taken in different environments and from several pieces with different characteristics in order to analyse all reconstruction results.

Table 2 shows a comparative study of the reconstruction results for the seven pieces analysed, taking into account different characteristics.

Table 2. Comparative table of results.

	Dragon's head	Helmet	Toy car	Computer mouse	Electric screwdriver	Cube of colours	Miniature Eiffel Tower
Colours	++	++	++	++	++	++	++
Textures	++	++	++	++	++	++	+
Model details	++	++	++	++	+	+	–
Holes definition	++	–	+	++	–	+	–
Borders	++	+	++	+	+	–	–

Particularly suitable: ++, *Moderately suitable*: +, *unsuitable*: –

When analysing the model of the dragon´s head, it can be observed that a very good quality was obtained for all defined aspects. It should be pointed out that the photographs were taken outdoors with natural light, achieving this way a more homogeneous lighting and a better definition of part details. In the case of the helmet, very good results were also obtained due to the photographs were taken outdoors, which affects the quality of the reconstruction. The only deficiency detected for this model was related with its holes, which were not completely defined. This is a major problem of this type of technology, requiring to be improved in future versions of the software. The toy car and the computer mouse were very similar in results, taking into account that dark or very bright areas can give surface resolution problems, so special care should be taken when photographing. Concerning the screwdriver, the obtained reconstruction presented a good surface finishing, but a lack of resolution between its contact surface and the support surface was observed, which could be solved minimizing the contact area by positioning the piece vertically. The cube of colours looked quite acceptable but, due to the flat faces of the piece, a little deficiency could be observed in the edges. Finally, the miniature Eiffel Tower reconstruction presented the worst results. To improve its results, a more neutral environment should have been used and higher quality photographs should have been taken in order to capture properly all the details.

The overall results are very acceptable in most pieces. The type of lighting and the surface characteristics of each piece can influence the reconstruction result, so a special attention is required during photographing (scanning phase).

5 Conclusions

Nowadays, reverse engineering is used in multiple professional fields that drive to develop different processes to obtain virtual models from real objects. This study assess the use of open source applications as data acquisition tools within a reverse engineering process. These applications can be used as an initiation to this process.

In the present study, a software-based alternative has been used to digitize seven different pieces, applying scanning and 3D reconstruction processes by using images, a method known as Image-Based 3D Modeling and Rendering (IB3DMR), in which a three-dimensional model can be generated from a set of 2D photographs. The chosen tool has been Autodesk ReCap and, as demonstrated, it offers a very good quality in most of the reconstructions obtained.

This type of methodology permits obtaining acceptable virtual objects to perform the simulations required in a determined design process, to analyse characteristics and surface properties of any kind of piece, etc. In addition, this type of tool does not have a great rigour in the precision of the equipment, but it emphasizes its ease of use, low cost and flexibility, all of them characteristics that make these tools very suitable in a wide variety of fields of application.

Funding. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Blais, F.: Review of 20 years of range sensor development. J. Electron. Imaging **13**(1), 231–243 (2004)
2. Remondino, F., El-hakim, S.: Image-based 3D modelling: a review. Photogramm. Rec. **21**(115), 269–291 (2006)

3. Cavas Martínez, F., et al.: Reconstrucción tridimensional de frutas a partir de imágenes digitales para su análisis. In: I Symposium Nacional de Ingeniería Hortícola “La Agromótica en la Horticultura”, Orihuela, España (2014)
4. Cavas Martínez, F., et al.: Empleo de imágenes digitales para reconstrucción tridimensional de componentes. In: 18th International Congress on Project Management and Engineering (AEIPRO 2014), Alcañiz, España (2014)
5. Cervera, O., et al.: Técnicas de modelado tridimensional y su aplicación en la auralización de espacios. In: VIII Congreso Ibero-Americano de Acústica, Évora, Portugal (2012)
6. Santagati, C., Inzerillo, L.: 123D catch: efficiency, accuracy, constraints and limitations in architectural heritage field. Int. J. Herit. Digit. Era **2**(2), 263–290 (2013)
7. Barsanti, S.G., Remondino, F., Visintini, D.: Photogrammetry and laser scanning for archaeological site 3D modeling - some critical issues. In: 2nd Workshop on the New Technologies for Aquileia, NTA 2012, Aquileia (2012)
8. Butnariu, S., Gîrbacia, F., Orman, A.: Methodology for 3D reconstruction of objects for teaching virtual restoration. Int. J. Comput. Sci. **3**(01), 16–21 (2013)
9. Opitz, R., et al.: Close-range photogrammetry vs. 3D scanning: comparing data capture, processing and model generation in the field and the lab. In: CAA 2012 (2012)
10. Venkatesh, S., Ganeshkar, S.V., Ajmera, S.: Image-based 3D modelling: a simple and economical technique to create 3-D models of the face. Int. J. Health Sci. Res. **2**(6), 93–99 (2012)
11. Rekittke, J., Paar, P., Ninsalam, Y.: Foot soldiers of geodesign. In: Peer Reviewed Proceedings of Digital Landscape Architecture 2012 at Anhalt University of Applied Sciences, pp. 199–210 (2012)
12. Santamaría, J., Sanz, F.: 3D Land modelling from photos. In: VII Congreso Ibérico de Agroingeniería y Ciencias Hortícolas (2013)
13. Lerma, J.L., et al.: Range-based versus automated markerless image-based techniques for rock art documentation. Photogramm. Rec. **29**(145), 30–48 (2014)
14. Lavy, A., et al.: A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. Methods Ecol. Evol. **6**(5), 521–531 (2015)
15. Parras, D., et al.: Empleo de una herramienta de bajo coste para la digitalización de piezas mecánicas. In: 20th International Congress on Project Management and Engineering, Cartagena (Murcia) (2016)

Simulation Sickness Evaluation While Using a Fully Autonomous Car in a Head Mounted Display Virtual Environment

Stanislava Rangelova[1,2](✉), Daniel Decker[1], Marc Eckel[1], and Elisabeth Andre[2]

[1] BMW AG, Munich, Germany
{stanislava.rangelova,daniel.decker,marc.eckel}@bmw.de
[2] University of Augsburg, Augsburg, Germany
andre@informatik.uni-augsburg.de

Abstract. Simulation sickness is a condition of physiological discomfort felt during or after exposure to a virtual environment. A virtual environment can be accessed through a head mounted display which provides the user with an entrance to the virtual world. The onset of simulation sickness is a main disadvantage of virtual reality (VR) systems. The proof-of-concept presented in this paper aims to provide new insights into development and evaluation of a VR driving simulation based on consumer electronics devices and a 3 Degrees-of-Freedom (3 DOF) motion platform. A small sample (n = 9) driving simulator pre-study with within-subjects design was conducted to explore simulation sickness outbreak, sense of presence and physiological responses induced by autonomous driving in a dynamic and static driving simulation. The preliminary findings show that users experienced no substantial simulation sickness while using an autonomous car when the VR simulation included a motion platform. This study is the basis for more extensive research in the future. Future studies will include more participants and investigate more factors that contribute to or mitigate the effects of simulation sickness.

Keywords: Simulation sickness · Virtual reality · Driving simulation
Consumer electronics · Head mounted display · Autonomous driving
Pilot study

1 Introduction

Immersive Virtual Reality (VR) driving environments have certain advantages when it comes to testing new automated driving concepts. Modern computation technology and sensors enable fully automated cars. Thus the interest in this type of driving has been increased. With VR systems a safe, fully controlled and still high fidelity environment can be provided at a fraction of the costs of real driving studies. In other words, an environment can be created which is close to a realistic driving experience without the liability issues or high costs [1]. However, one of the disadvantages of the VR systems is simulation sickness. Simulation sickness is a form of motion sickness which is induced by virtual environments and is also referred to as VR sickness or cybersickness [2].

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 155–167, 2018.
https://doi.org/10.1007/978-3-319-91581-4_12

Simulation sickness is a condition of physiological discomfort felt during or after exposure to a virtual environment. According to cue conflict theory, the discrepancy between visual and motion cues is one of the assumed reasons for simulation sickness while using driving simulators [3]. A motion system might be able to replace the missing motion cues in the VR driving environment. In this paper, a pilot study is presented to assess the feasibility of a future experiment to evaluate simulation sickness in an autonomous VR driving environment. Automated driving is categorized into five levels. Fully automated cars have the highest level of automation and the driving system has full control over all driving tasks under all road conditions which are managed by a human driver in lowest level [4]. In particular, the effect of adding a motion platform to such a system is of interest. The study has two objectives: First, testing the experimental setup and the data collection methods and second, gathering preliminary results on the participants' responses. The data was collected through subjective questionnaires and interviews, and objective physiological measurements.

In the next section a brief overview of virtual reality, simulation sickness, and related work will be provided. The methodology for the user study and the experimental setup will be described in Sect. 3, followed by the results of the trials. This paper concludes with a discussion of the results and shows a way towards future research.

2 Background

2.1 Virtual Reality

VR is a technology that enables a person to experience a computer-simulated environment and interact with it. These virtual worlds can be accessed through a Head-mounted Display (HMD) which provides access to these virtual environments. VR is a well-known technology since the 1960s [5]. Despite its half-century long existence, VR was neither widely adopted by consumers nor the industry. However, in recent years VR was rediscovered as a potential gaming and visualization platform. The year 2012 was a turning point for the technology when a Kickstarter project called Oculus Rift [6] released a user-friendly, low-cost HMD. This development expanded the interest and adoption of VR technology exponentially [7]. Oculus Rift was developed with the intention to serve the consumer electronics market. Despite being a low-cost device, the Rift offers key advantages over previous, much more expensive HMDs that have been used by the industry; a wide Field-Of-View (FOV), high-resolution display panels, affordable head-tracking technology, and a combination of hard- and software to ensure a minimal latency [8]. Besides gaming, VR technology is widely adopted by industry and academia in fields such as architecture, health, aerospace, and military. In the automotive sector, one of the applications for VR is driving simulation. It can be used for testing new interior concepts and interfaces in a realistic three-dimensional driving environment.

2.2 Simulation Sickness

Cue conflict theory is also known as sensory conflict, neural mismatch, and sensory rearrangement theory. It is the most widely adopted theory of the origins of simulation sickness. It was originally developed to explain motion sickness, later it was discovered that the theory is also applicable to simulation sickness. In (simulated) motion conditions a mismatch between visual, vestibular and muscular proprioceptive systems can arise [9]. The brain is confused by the received information and this mismatch can result in an immediate response. This can lead to physical discomfort like disorientation, nausea or eyestrain. For example, if a person is using a static driving simulator, the vestibular system indicates that the body does not move. The visual system sends opposing signals. Due to this cue conflict, the brain concludes that normal functions of the body are interrupted, in extreme circumstances such a mismatch can lead to emesis.

Poison theory, also known as evolutionary theory, suggests that unnatural movement can cause physical discomfort. The human body learned naturally how to move in the surrounding environment [10]. If a person is in a situation, which involves uncommon movements like walking in a virtual environment, the body misreads the sensory input information. Sensory systems send signals to the brain to bring awareness of the intake toxin. As a result, the person can feel disorientation, overall discomfort and ultimately an emesis response to dispense of the suspected toxins.

Decades later, Stoffregen disagreed with the cue conflict theory and suggested another theory on why people get sick known as the postural instability theory. The theory is based on the fact that the main purpose of the human body is to maintain postural stability, and when this balance is interrupted, the person feels discomfort (e.g. disorientation, nausea, and dizziness) [11].

The widest used measurement to quantify the level of sickness is the simulation sickness questionnaire (SSQ) [12]. The questionnaire consists of 16 questions where each question has four possible answers e.g. none, slight, average and severe. According to the SSQ scoring system, each item falls into one of three clusters: nausea, disorientation and oculomotor. Other measurements include physiological signals, such as electro-dermal activity (EDA) and electrocardiogram (ECG). Skin conductance levels, acquired by EDA, provides information about stress level which is related to simulation sickness outbreak [13]. The heart rate (HR), acquired by ECG, was related to simulation sickness onset in previous research [14].

2.3 Related Work

Research in VR driving simulation and, more specifically, driving simulation with HMDs has grown in the last years [15, 16]. There are a few studies focusing on VR driving simulator evaluation with HMD [1]. In a comprehensive study of simulation sickness in a virtual environment, Kolasinski [17] found that 42 factors are related to sickness outbreak. Factors, such as gender [18, 19], motion sickness history [20, 21], calibration [22], and latency [23, 24], have been investigated over the years. As there are many factors that contribute to simulation sickness, possible mitigation techniques are diverse and should be adapted to each specific use case. Regarding simulation sickness during simulated driving, research suggests that missing motion cues are one of the biggest problems of static driving simulators [25]. Missing motion cues are not the only contributing factors to simulation sickness. There is evidence that the low resolution HMDs in previous studies did not suffice to generate appropriate visual illusion [26]. This can contribute to simulation sickness and bring symptoms, such as blurred vision or disorientation. Hence, the visual quality of the simulation should be high to prevent these symptoms.

In a study comparing static and dynamic driving simulators, users who drove the motion driving simulator experienced fewer side effects. The results showed significantly reduced values of nausea, dizziness, eyestrain, and tiredness [25]. In a study, which compared a driver and a passenger exposed to the same motion simulation, it was reported that the participants in the driver's group experienced less motion sickness. The authors of the study conclude that the results are caused by the driver's concentration on the task [27]. These results suggest that users who ride a fully autonomous car will experience a higher level of sickness.

A proof-of-concept, presented in the next section, extends the use of VR driving environment for testing with addition of a fully autonomous driving environment. This concept could find application in future experiments in the domain of autonomous driving.

3 Methodology

3.1 Experimental Design and Procedure

A fully autonomous driving study with a within-subjects design was conducted. The participants were exposed to a scenario which did not require any intervention or monitoring from the user. Two conditions were presented: First, a scenario without motion platform and a scenario with additional motion. Each participant took part in both conditions with a 24 h gap between the sessions. Approximately 80% of the participants started with the static condition. The total duration of the driving simulation was around 11 min. The participants were instructed to sit comfortably on the driver's seat and explore freely the interior of the car or the surroundings. The simulation contained a traffic environment with no other movable visual assets (i.e. no traffic situations with vehicles or pedestrians). The driving simulation started with a right turn to a terminal of an airport. There the car stopped for 5 s and afterwards continued towards a highway. Before reaching the highway, the participant experienced driving down under a bridge. During the highway driving, the car changed the lane from the left to the right side. After that, the participants experienced driving up and down a hill, followed by a right turn to a country road including two left and two right turns. The simulation stopped when the car reached a specific point where the participant could see a city in the far distance.

3.2 Virtual Reality Driving Environment

For a realistic VR simulation environment which supports the current consumer version HMDs, a game development environment is required. For the proof-of-concept reported in this paper, a game engine named Unreal Engine 4 was employed. This game engine is widely used by game developers, and it delivers high quality graphics that can be optimized for VR applications. High quality graphics are desirable elements in the development of VR driving simulation [28]. Unreal Engine 4 was used in the following study for creating and executing the driving scene and the highest level of automated driving simulation (i.e. automated level 5). The VR driving system consisted of an HMD with a tracking system, a motion platform with four pneumatic actuators, and a computer for rendering the driving scene (Fig. 1). As the HMD, the Oculus Rift CV1 with a FOV of 110° and a resolution of 2160 × 1200 pixels per eye was used.

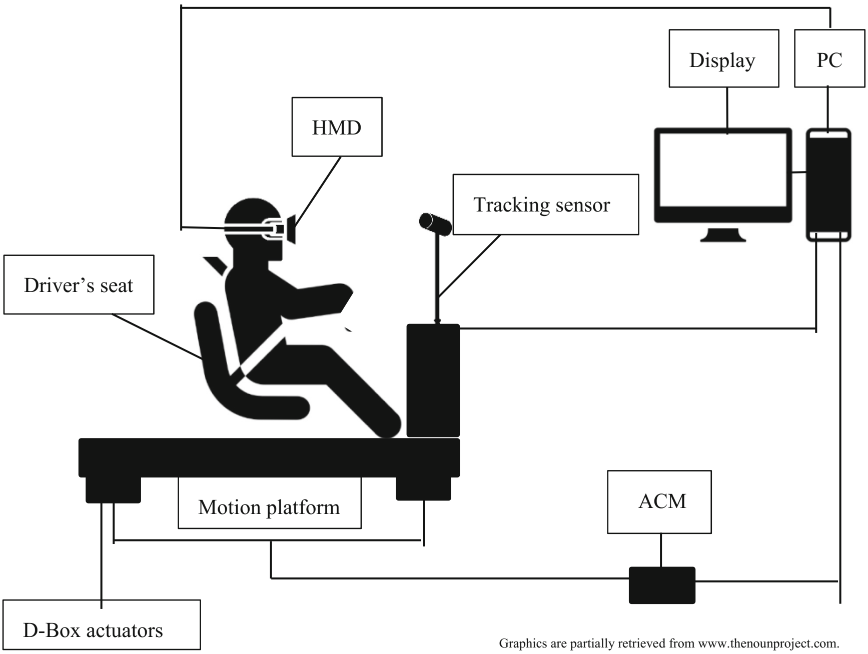


Fig. 1. The concept of VR driving simulation with HMD.

With a three Degrees of Freedom (3 DOF) the motion platform can simulate the most common movements during driving, namely, pitch, roll, and heave [29]. Each actuator has a maximum payload of 227 kg, a maximum acceleration of 1 g-force, a maximum angle of 15°, and a stroke of 152.4 mm. The four actuators are attached to the corners under the platform. Two actuators have one master box. Each box translates the signals from the Actuator Control Module (ACM) and controls the actuators. The ACM is connected to a high-end personal computer (PC), where a Motion code is executed. On the same PC, a 3D scene of the driving simulation is displayed. The tracking sensor was attached to the platform and therefore follows the platform's movements. This ensured that the platform's movements are not interpreted as head movements in the virtual scene. To create an immersive environment, a car seat was added to the platform for a more realistic experience. During the experiment, the 3D scene was shown on the display and the HMD (Fig. 2). This allowed the experiment supervisor to monitor the participants' progress.

Fig. 2. A screenshot shows part of the VR environment used for the concept testing.

3.3 Participants

Nine participants aged between 30 and 53 years (M = 35.67, SD = 7.04) took a part in the study. Only one female participated and therefore, gender as a variable was unequally distributed and could not be evaluated. Five of the participants described themselves as frequent drivers who drove more than 10 000 km in the past year. Regarding previous experience with driving simulation, 67% responded positively. Only one of the participants reported discomfort after being exposed for over 30 min to a driving simulator. Two participants reported that they play video games on a daily basis. In respect to previous experience with HMDs, 67% had used an HMD before this trial, one participant did not respond and one experienced simulation sickness during previous VR session. They had spent between 5 and 300 min (M = 73.3, SD = 111.8) in VR environments.

3.4 Measurements

For this pilot study, the objective physiological measurements ECG, EDA and subjective measurements, such as questionnaires were recorded. For measuring the physiological signals, 3-channel-ECG and skin conductance level data were recorded with medical sensors by g.Tec Medical Engineering GmbH with a sampling frequency of 512 Hz [30]. A baseline signal was recorded for 2 min in a resting position in the VR environment prior to the driving scenario.

Two subjective questionnaires were handed out before and two after each simulation. Prior to the trial, the participants were given a questionnaire asking basic demographic and biographical data. Also before the trial, a short version of the motion sickness susceptibility questionnaire (MSSQ) was used [31]. The post-questionnaires consisted of the SSQ [12] and the iGroup Presence Questionnaire (IPQ). A few questions regarding enjoyment [32] were included in the IPQ questionnaire to measure the users' emotional reaction to the VR simulation. Additionally, interviews were developed to collect participants' responses. The aim of these interviews was to acquire additional information about the trial.

3.5 Data Analysis

The study was designed as a pilot study to assess the feasibility and possible procedure for a future study. Due to the small sample size, the data was analyzed with the help of descriptive statistics. The HR was obtained from ECG signals [14]. Skin conductance level (SCL) was obtained from EDA signals and processed with the Ledalab software [33].

4 Results

The results of the subjective data collected through SSQ indicated that no clear trend could be observed between the motion and static conditions regarding simulation sickness. Only light to non-existent symptoms in both conditions were reported. The overall score indicated that the *general discomfort* and *stomach awareness* score was higher in the static condition. However, *difficulty focusing* and *fullness of the head* shows a trend towards higher scores in the motion platform condition. Figure 3 shows the calculated clusters' scores based on the SSQ's weight system. The symptoms of nausea cluster which showed changes in the responses are *sweating*, *salvation increased* and *stomach awareness*. In the static condition, one participant had a slight *sweating* and three participants had slight *stomach awareness*. In the motion platform condition, *sweating* and *stomach awareness* were not reported and one participant felt *increased salivation*.

The results of the MSSQ-short questionnaire showed that two users had a high level, two had moderate, and five had a low level of susceptibility to motion sickness. From the users with high level, one showed a lower SSQ score in the static condition. One of the users with moderate level demonstrated a lower SSQ score in the static, while the third showed a lower SSQ score in the motion platform condition. Regarding the sense of presence, the answers to the question "I was involved in the virtual reality experience." from IPQ indicated a tendency towards the motion platform condition. The question "I experienced a delay between my actions and expected outcome." suggested that the static condition provoked a higher experienced delay compared to the motion platform condition. The last four questions regarding enjoyment showed mostly positive answers. No differences in HR and SCL could be observed (Figs. 4 and 5).

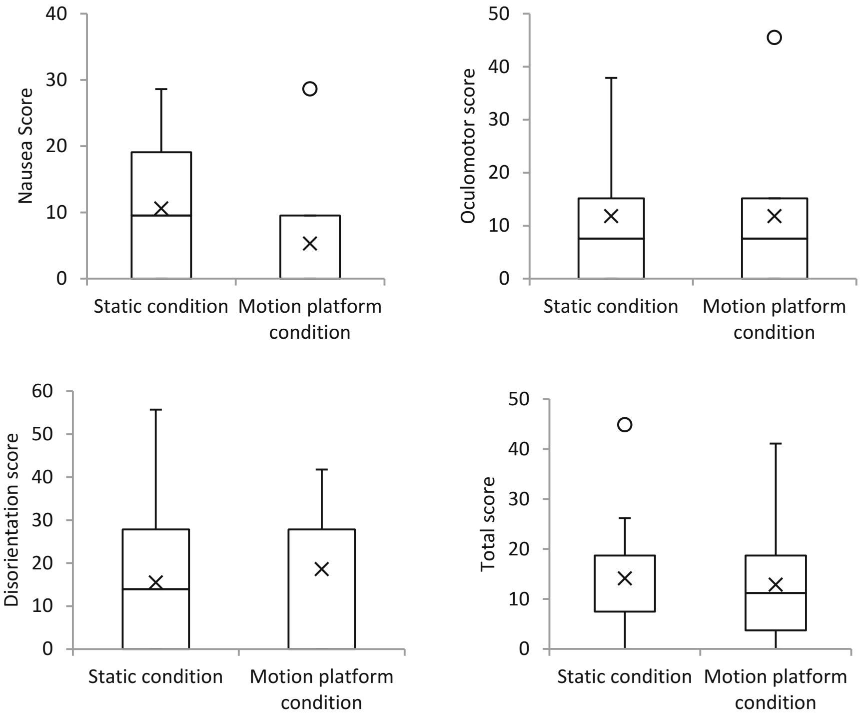


Fig. 3. SSQ score divided into clusters for static and motion platform conditions.

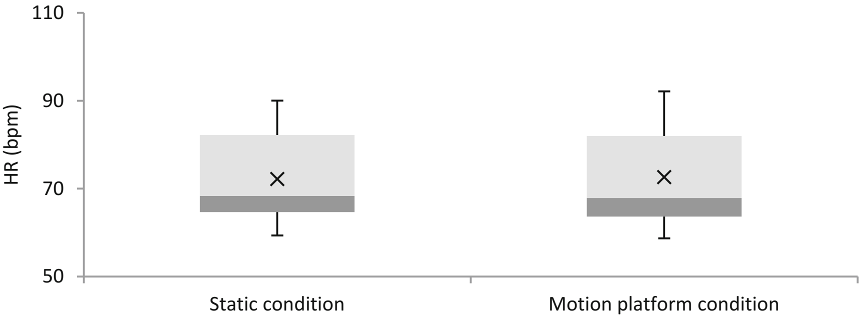


Fig. 4. Mean HR in static and motion platform conditions.

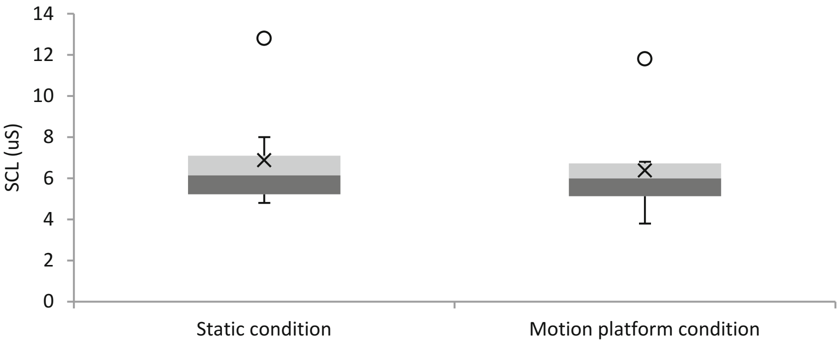


Fig. 5. Mean SCL in static and motion platform conditions.

The participants were asked to describe their overall experience from the VR autonomous driving. The experience was described as interesting and exciting, as one participant said: "It was very exciting because it was my first time to try out a driving simulation with virtual reality and at the same time it was a bit strange." Another participant, responding to the question when the highest amount of discomfort was felt, said: "Breaking and acceleration in both conditions made you feel strange. The motion felt mostly like a vibration." Other participants responded to the same question that the turns were the most unpleasant part of the simulation. One participant stated that "the simulation felt very artificial like nothing is happening there which got boring at some moment." And another commented "It was much better than using only a screen." One of the participants felt a light discomfort shortly after the trial which increased with time. One hour later, the participant got nauseous. According to the experiment's notes, a second participant experienced similar discomfort, but after a short duration of time returned to a normal level. The overall response to the VR autonomous driving experience was positive and participants expressed the willingness to participate again.

A more detailed information for a particular participant is shown in Fig. 6. It shows that the HR and SCL were higher in the static condition. The last event, *car drives up,* has lower HR level than the previous event, *car drives down*, in the motion platform condition. This differs from the static condition where the same event had a higher HR level than the *car drives down* event.

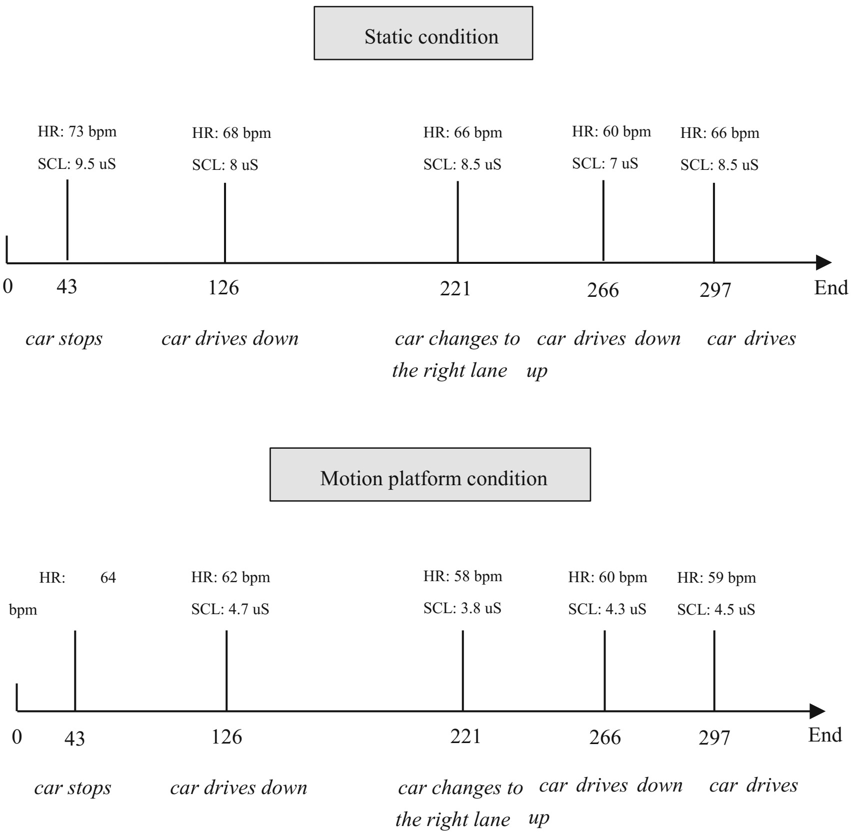


Fig. 6. HR and SCL over time (in seconds) in static and motion platform condition during the following events: *car stops*, *car drives down*, *car changes to the right lane*, *car drives down*, and *car drives up*.

5 Discussion and Future Work

A pilot study was conducted to test a proof-of-concept and collect preliminary results of the effect of autonomous VR driving simulation on simulation sickness. The findings have shown that a setup with HMD could be used for evaluating new interior concepts in an autonomous environment. No substantial simulation sickness was reported while using an autonomous car in a VR simulation whether a motion platform was used or not. A possible explanation for these findings is the low speed which was used. The speed was reduced in order to minimize the simulation sickness provocation.

Most of the participants had previous experience with virtual environments and were accustomed to wear HMDs repeatedly. Hence, the used sample is not representative of the general population. Literature suggests that a person has the ability to adapt to a virtual environment to some extent which can lead to a decrease of motion sickness symptoms [34]. Some of the participants felt less sick in the static condition. These findings may be due to the response time of the platform which was too high and therefore, the induced motion stimuli might be wrong. This could have made simulation sickness outbreak worse.

An interesting finding is that the frequency of *general discomfort*, *difficulty focusing*, and *dizziness with eyes open* was reduced in the motion platform condition. A possible reason for that could be that motion cues contributed to the overall known driving behavior which leads to a feeling of comfort while using an autonomous car. Intriguingly enough, the frequency of *sweating* and *stomach awareness* changed to zero in the motion platform condition. Sweating, and more specifically cold sweating is the body's reaction to stress [35]. This indicates that participants in motion platform condition experienced less stress. However, this condition also induced fatigue and fullness of the head. The HR and SCL data differed in the static condition regarding the event *car drives up* in relation to the previous event *car drives down*. A possible explanation is that in the motion platform condition the discrepancy between the visual and motion cues is less during the *car drives up* event. However, this participant showed higher SSQ total score in motion platform condition. A possible reason for that could be factors such as stress from work, level of tiredness, and general well-being. These factors could not be easily controlled but could be taken into consideration in future experiments. No differences in HR and SCL could be observed. This might be caused limitations of the experimental design, a small sample size and an order of the scenarios that were not counterbalanced.

One of the reasons given by the participant for simulation sickness outbreak was the turning during the simulation. The left and right turns were felt sometimes unpleasant to the users especially in the static condition. One possible explanation is that during the steering of the automated car the discrepancy is higher which corresponds to cue conflict theory [36]. However, in the motion platform condition, discomfort might have occurred because the user's driving style did not match completely with the style of the autonomous driving. The aftereffects which were experienced by two of the participants also have been reported in other studies [37]. The assumption that the higher SSQ score is associated with higher skin conductance level could not be observed in the sample.

The limitations of this study are the small sample size, unequal distribution across gender and age, the design was not counterbalanced, and there was no familiarization scenario. Also, subjects participated on two separate days, which might add more confounding factors like different daytime or different well-being state.

Considering the findings from the concept design, a follow-up study might include a different tilting angle and vibration frequency of the motion platform. That way, different motion levels, which would represent different speed, could be investigated. The rare occurrence of simulation sickness in the pre-study points to a larger sample size, a better balanced order of testing, and a longer driving time in future experiments. Furthermore, testing various environments might be further investigated. This study strengthens the idea that a VR environment might be used for experiments in the autonomous driving domain. An immersive environment can bring flexibility for testing concepts which are coming in the near future.

References

1. Walch, M., et al.: Evaluating VR driving simulation from a player experience perspective, pp. 17–25 (2017)
2. Mousavi, M., Jen, Y.H., Musa, S.N.B.: A review on cybersickness and usability in virtual environments. In: Advanced Engineering Forum, vol. 10, no. November 2016, pp. 34–39 (2013)
3. Brooks, J.O., et al.: Simulator sickness during driving simulation studies. Accid. Anal. Prev. **42**(3), 788–796 (2010)
4. SAE International: Summary of SAE international levels of driving automation for on-road vehicle (2014)
5. Sutherland, I.E.: The ultimate display (1965)
6. Kickstarter: Oculus Rift: step into the game by Oculus—Kickstarter (2012). https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game?lang=de. Accessed 29 Jan 2018
7. Reede, E., Bailiff, L.: When virtual reality meets education—TechCrunch (2016). https://techcrunch.com/2016/01/23/when-virtual-reality-meets-education/. Accessed 24 Jan 2018
8. Carmack, J.: Latency mitigation strategies (2013). https://web.archive.org/web/20140719053303/, http://www.altdev.co/2013/02/22/latency-mitigation-strategies/. Accessed 29 Jan 2018
9. Crampton, G.: Motion and Space Sickness. CRC Press, Boca Raton (1990)
10. Treisman, M.: Motion sickness: an evolutionary hypothesis. Science **197**(4302), 493–495 (1977). (New York, N.Y.)
11. Riccio, G.E., Stoffregen, T.A.: An ecological theory of motion sickness and postural instability an ecological theory of motion sickness and postural instability. Ecol. Psychol. **3**(3), 195–240 (1991)
12. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**(3), 203–220 (1993)
13. Min, B.-C., Chung, S.-C., Min, Y.-K., Sakamoto, K.: Psychophysiological evaluation of simulator sickness evoked by a graphic simulator. Appl. Ergon. **35**(6), 549–556 (2004)
14. Bruck, S., Watters, P.A.: The factor structure of cybersickness. Displays **32**(4), 153–158 (2011)
15. Coates, N., Ehrette, M., Hayes, T., Blackham, G., Heidet, A., Kemeny, A.: Head-mounted display in driving simulation applications in CARDS. In: Proceedings of Driving Simulator Conference, vol. 3, no. 1, pp. 33–43 (2002)
16. Kemeny, A.: From driving simulation to virtual reality. In: Proceedings of 2014 Virtual Reality International Conference on VRIC 2014, pp. 1–5 (2014)
17. Kolasinski, E.M.: Simulator sickness in virtual environments. United States Army Res. Inst. Behav. Soc. Sci. **1027**(4), 68 (1995)
18. Munafo, J., Diedrick, M., Stoffregen, T.A.: The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. Exp. Brain Res. **235**(3), 889–901 (2017). https://doi.org/10.1007/s00221-016-4846-7
19. Graeber, D.A., Stanney, K.M.: Gender differences in visually induced motion sickness. In: Proceedings of Human Factors and Ergonomics Society Annual Meeting, vol. 46, no. 26, pp. 2109–2113 (2002)
20. Wright, R.H.: Helicopter simulator sickness: a state-of-the-art review of its incidence, causes, and treatment (1995)

21. Matas, N.A., Nettelbeck, T., Burns, N.R.: Dropout during a driving simulator study: a survival analysis. J. Saf. Res. **55**, 159–169 (2015)
22. Rebenitsch, L., Owen, C.: Review on cybersickness in applications and visual displays. Virtual Real. **20**(2), 101–125 (2016)
23. Davis, S., Nesbitt, K., Nalivaiko, E.: A systematic review of cybersickness. In: Proceedings of 2014 Conference on Interactive Entertainment, pp. 1–9. ACM, Newcastle, NSW, Australia (2014)
24. Bailey, R.E., Arthur Iii, J.J., Williams, S.P.: Latency requirements for head-worn display S/EVS applications, vol. 5424, pp. 98–109 (2004)
25. Aykent, B., Merienne, F., Guillet, C., Paillot, D., Kemeny, A.: Motion sickness evaluation and comparison for a static driving simulator and a dynamic driving simulator. J. Automob. Eng. **288**(7), 818–829 (2014)
26. Settgast, V., Pirker, J., Lontschar, S., Maggale, S., Gütl, C.: Evaluating experiences in different virtual reality setups. In: Wallner, G., Kriglstein, S., Hlavacs, H., Malaka, R., Lugmayr, A., Yang, H.-S. (eds.) ICEC 2016. LNCS, vol. 9926, pp. 115–125. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46100-7_10
27. Rolnick, A., Lubow, R.E.: Why is the driver rarely motion sick? The role of controllability in motion sickness. Ergonomics **34**(7), 867–879 (1991)
28. Mourant, R.R., Schultheis, M.T.: A HMD based virtual reality driving simulator. In: First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design A, pp. 300–304 (2001)
29. Arioui, H., Hima, S., Nehaoua, L.: 2 DOF low cost platform for driving simulator: modeling and control. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 1206–1211 (2009)
30. g.Tec: g.Tec Medical Engineering
31. Golding, J.F.: Scoring the MSSQ-Short, pp. 2–3 (2007)
32. Lin, J.J.-W., Duh, H.B.L., Parker, D.E., Abi-Rached, H., Furness, T.A.: Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In: Proceedings of IEEE Virtual Reality, vol. 2002, no. February 2002, pp. 164–171 (2002)
33. Benedek, M., Kaernbach, C.: A continuous measure of phasic electrodermal activity. J. Neurosci. Methods **190**(1), 80–91 (2010)
34. Lewis, C.H., Griffin, M.J.: Human factors consideration in clinical applications of virtual reality. Stud. Health Technol. Inform. **44**, 35–56 (1997)
35. Meehan, M., Razzaque, S., Insko, B., Whitton, M., Brooks, F.P.: Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. Appl. Psychophysiol. Biofeedback **30**(3), 239–258 (2005)
36. Reason, J., Brand, J.J.: Motion sickness (1975)
37. Llorach, G., Evans, A., Blat, J.: Simulator sickness and presence using HMDs: comparing use of a game controller and a position estimation system. In: Proceedings of 20th ACM Symposium on Virtual Reality Software and Technology, pp. 137–140. ACM, Edinburgh (2014)

Visualizing Software Architectures in Virtual Reality with an Island Metaphor

Andreas Schreiber[1](✉) and Martin Misiak[2]

[1] Intelligent and Distributed Systems, German Aerospace Center (DLR), Linder Höhe, 51147 Köln, Germany
Andreas.Schreiber@dlr.de
[2] TH Köln – University of Applied Sciences, Betzdorfer Straße 2, 50679 Köln, Germany
martin.misiak@th-koeln.de

Abstract. Software architecture is abstract and intangible. Tools for visualizing software architecture can help to comprehend the implemented architecture but they need an effective and feasible visual metaphor, which maps all relevant aspects of a software architecture and fits all types of software. We focus on the visualization of module-based software—such as OSGi, which underlies many large software systems—in virtual reality, since this offers a much higher comprehension potential compared to classical 3D visualizations. Particularly, we present an approach for visualizing OSGi-based software architectures in virtual reality based on an island metaphor. The software modules are visualized as islands on a water surface. The island system is displayed in the confines of a virtual table where users can explore the software visualization on multiple levels of granularity by performing intuitive navigational tasks. Our approach allows users to get a first overview about the complexity of the software system by interactively exploring its modules as well as the dependencies between them.

Keywords: Virtual reality · Software visualization · Software analysis

1 Introduction

Software is abstract and intangible. With increasing functionality, its complexity grows and hinders its further development. Visualization techniques, that map intangible software aspects onto visually perceivable entities, help to enhance the understandability and reduce the development costs of software systems [10]. Over the years a number of two and three dimensional visualization approaches have been proposed. However, the visualization of software in *virtual reality (VR)* still remains a sparsely researched field. While it offers a much higher comprehension potential compared to classical three dimensional visualizations, it also requires a different approach, as the requirements on a usable VR application

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 168–182, 2018.
https://doi.org/10.1007/978-3-319-91581-4_13

are much higher than those of a classical desktop application. This is derived from the substantially higher immersion degree, that this medium can achieve.

We present an approach for visualizing OSGi-based software projects in virtual reality. OSGi (Open Services Gateway Initiative), is a module-based, service-oriented framework specification for Java. OSGi is widely used in the *Eclipse* ecosystem. It centers on application development using modular units, called bundles. A bundle is a self-contained unit of classes and packages, which can be selectively made available to other bundles. The OSGi service layer handles communication between service components, which can reference or implement specific interfaces [18]. Popular implementations of the OSGi specification are *Apache Felix*, *Equinox*, or *Knopflerfish*.

Our goal is to provide a high level overview of the underlying software architecture to the user, while minimizing the experienced simulator sickness and enabling an intuitive navigation. Although our implementation targets OSGi-based software, the presented concepts are applicable to a large variety of use cases, such as other module-based software architectures.

Our main contributions are:

– We present a novel real-world metaphor based on an island system for visualizing module-based software architectures (Sect. 3).
– We present our approach for the exploration of software projects in VR, with focus on high level comprehension and improvements to user comfort (Sect. 4).

2 Software Visualization

Software visualization is a very large research field. Existing work can be classified based on multiple categories. A differentiation between static and dynamic aspects of a software can be made. While dynamic aspects capture information of a particular program run, static aspects are valid for all execution paths of a software. Additionally, they can be extended to capture the entire evolution of a software architecture.

Software visualization can be made on roughly three different levels of abstraction [7]. While the lowest abstraction level deals with the source code, the highest abstraction level deals with the entirety of the software architecture and belongs to the most important in software visualization [1]. They convey the underlying hierarchical component structure, the relationships between these components and the visual representation usually contains some form of code quality metrics.

A software visualization can consist of one or more views. Each view can use its own visualization approach and can therefore focus on different aspects of the software. Multi-view, as opposed to single-view approaches, can represent a broad range of information of varying granularity levels, however at the cost of imposing a significant cognitive burden and making a communication on common grounds between users more difficult [17].

Visualizations can be made in the two dimensional and three dimensional space. While 2D visualizations are easier to navigate and interact with, they do not scale particularly well for large data sets. To avoid a cluttered view, 2D visualizations rely on multiple views. 3D approaches resort to the added dimension to increase the information density of the visualization, without exposing the user to any additional cognitive load, as the task of processing 3D objects is completely shifted to the perceptual system [13]. They are more effective at identifying substructures and relationships between objects [5,22] and represent real-world metaphors more closely.

3 Visual Metaphors and Environment

3.1 Islands

After careful analysis of the requirements the visualization had on the metaphor, the concept of mapping the software system onto an island metaphor was chosen. The entire software system is represented as an ocean with many islands on it. Each island represents an OSGi bundle and is split into multiple regions. Each region represents a Java package and contains multiple buildings. The buildings are the representatives of the individual class types which reside inside of a package. Each region provides enough space to accommodate all of its buildings without overlapping, and hence the overall size of an island is proportional to the number of class types inside of a bundle.

The island metaphor provides a hierarchical structure with three different levels (island, region, and building). The navigation between these layers should be based on our natural understanding of spatial relationships and be therefore dependent on the relative size of the elements in the users view frustum. Hence, the transition between the levels happens implicitly, as the user moves closer or further away from an element, or the element itself is scaled. This avoids the introduction of additional complexity into the navigation.

The metaphor is flexible enough to be extended for more than three abstraction levels. Individual island groups can form archipelagos, which would provide an additional abstraction level. In the opposite direction, each island region could be interpreted as a country, which would open up even more possible hierarchical subdivisions.

The islands metaphor has several advantages for software visualization. Islands express the aspect of decoupled entities, coexisting in the same environment very clearly, which makes them a good candidate for representing software modules. Additionally, islands can be relocated at run-time, while maintaining a certain plausibility. A software evolution visualization could benefit from this property, as the island movements would reflect the dependency changes within the system.

The islands of our visualization metaphor are aimed at having a high resemblance to their real-world counterparts and in thus, emphasizing the plausibility of the metaphor. Each region represents a package and has an irregular, rugged shape, similar to countries when seen on a map. These regions share borders, and together, they determine the shape of the island (Figs. 1 and 2). The individual cells of a region are designed to provide enough accommodation area

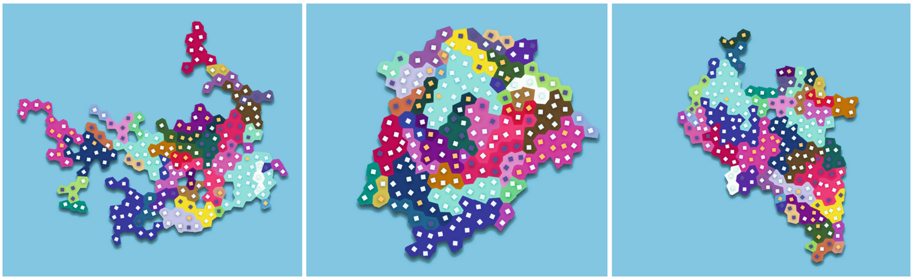


Fig. 1. (left) A minimal cohesion factor leads to very rugged islands with many holes. (middle) A very high cohesion factor reduces holes greatly and creates compact islands. (right) Our dynamic cohesion factor, combined with the claiming of large regions first, the island preserves some of the ruggedness, yet it minimizes holes.

Fig. 2. A range of different coast shapes, created with specific height profiles.

for buildings to be placed on top. To maximize their perceivability from afar a multi-storey building representation is chosen. For the implemented prototype, a *Lines of Code* metric was chosen, where for every n lines of code, a storey is added to the building.

Island Construction. The island construction is based on claiming cells in a Voronoi diagram (analogous to the work of Yang and Biuk-Aghai [23], while we use a Voronoi diagram instead of a hexagonal grid as the underlying tile structure). Additionally, no hierarchical claiming is performed, as all regions of an island are considered equivalent. This reflects the way packages are interpreted by Java, as the hierarchical naming convention is only relevant from a developers perspective.

The first step in the construction is to create a Voronoi diagram from a point distribution. The most aesthetically pleasing islands were achieved with points exhibiting a blue noise characteristic. In the next step, each package claims multiple cells of the created Voronoi diagram, corresponding to the number of contained classes. Cells are claimed one at a time and only cells next to already existing entities can be claimed.

To create rugged and irregular shapes for the package representations, the cells are selected probabilistically. To avoid non-continuous regions induced by a random selection mechanism, we employ an estimating function as described by Yang and Biuk-Aghai [23].

Before a new tile is selected, each eligible cell counts its already claimed neighbors. If a cell is surrounded with n claimed neighbors, the probability of it being a hole grows with n. A score S_n is calculated for each candidate, based on

$$S_n = b^n \tag{1}$$

where b is a user definable cohesion factor. Once the scores are known, a new cell can be selected, where the probability of each candidate is directly proportional to its score S_n. Higher b values result in less holes, but also more regular and compact shapes.

To preserve the rugged appearance of an island, we use a simple extension of the cohesion factor. Defining b_{min} and b_{max}, the cohesion factor can be varied on a per region basis, depending on their size. While the smallest region is assigned b_{min}, the cohesion factor is interpolated towards b_{max} for larger regions. Additionally, the regions are claimed in descending order, starting with the largest package first. This results in islands which contain smaller, irregular regions at their edge, while the larger, more regular regions reside in the interior (Fig. 1 right). From a usability perspective, this layout is more advantageous for VR based interaction, as smaller regions are harder to select when surrounded by larger ones.

Once all packages have claimed their cells, the coast area can be added. This is done by claiming neighboring cells of the existing island boundary. Each time the boundary is expanded outwards a new height value is associated with its cells. A user defined height profile controls this process (Fig. 2), where each entry expands the coast by one cell and assigns the stored height value. In the final construction step, a polygonal mesh is generated from all claimed cells in the Voronoi diagram using triangulation.

3.2 Visualization of Dependencies

Dependencies Between Modules. Due to the architecture oriented focus of the presented software visualization, the dependencies between individual modules are of high importance. Building on the island metaphor, an import and export port is added to each island. These ports are situated along the coast line and manage the incoming and outgoing dependencies. To visualize them, two orthogonal types of approaches are considered. An explicit and an implicit dependency visualization.

Explicit Visualization. Building on the simplicity of straight lines, import and export arrows (Fig. 3) are used to explicitly visualize the package dependencies

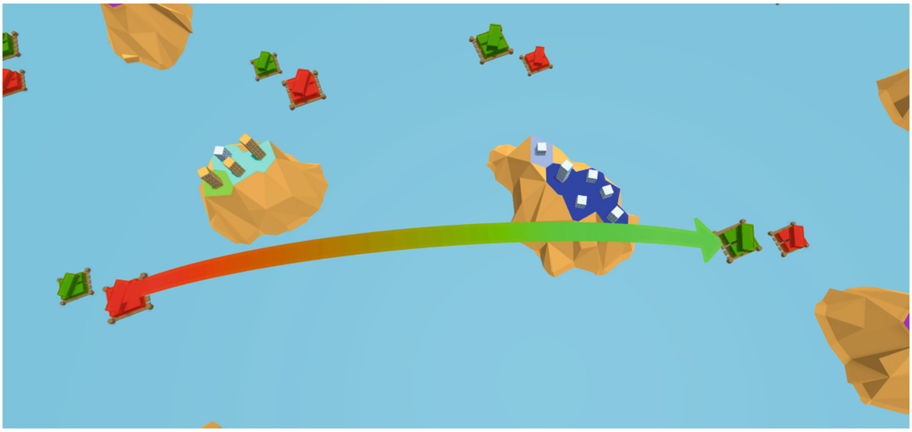


Fig. 3. Explicit dependency visualization via an arced arrow. The island on the right imports packages from the bundle on the left.

between bundles. In the geographic context, such arrows are encountered in flow maps [11] and visualize the movement of various resources or entities, from one point to another, while the arrow width is proportional to the moved volume. The resulting dependency visualization is similar to a discrete flow map, as implemented by Tobler [19]. To reduce the intersection problem of straight lines, the arrows follow a vertical arc. The start and end points are at the height of a port, while towards the middle segment the height increases, reaching its maximum halfway between the anchor points. The arrows maintain throughout a constant curvature. As a result, longer arrows also span a greater height range. A color gradient, together with the arrow head indicate the dependency direction. The width is mapped to the number of packages which are being imported or exported over the given connection.

Implicit Visualization. We use the island adjacency to implicitly represent the dependency strength (Fig. 4). The island layout is computed with the help of an iterative, force-directed graph layout algorithm, based on the work of Eades [3]. In it, nodes are interpreted as particles, which are influenced by attractive and repulsive forces from other particles. These forces are accumulated and applied to each particle at the end of the iteration. Attractive forces are exerted between nodes, which are connected by an edge. The force is dependent on the distance d between the two nodes and the variables c_1 and c_2.

$$F_a = c_1 \cdot log(d/c_2) \tag{2}$$

F_a can be interpreted as a spring like force defined by the stiffness factor c_1 and the unloaded spring length c_2. In contrast to Eades, who focused on layouts with uniform edge lengths, we compute c_2 on a per edge basis to reflect the relative dependency strength between the nodes in question as

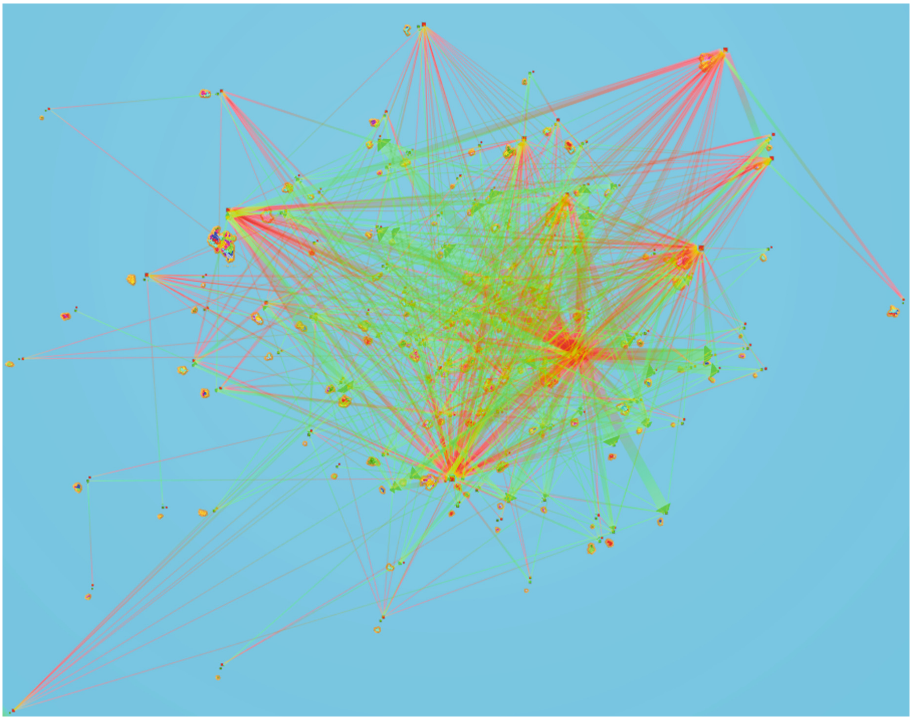


Fig. 4. Island placement based on the presented force-directed layout algorithm. Islands with the highest dependencies are accumulated in the middle, while independent islands are pushed outwards.

$$c_2 = c_3 \cdot \frac{i_{max}}{i_A + i_B}, \quad (3)$$

where i_{max} is the project wide largest number of bidirectionally imported packages per edge, i_A is the number of packages bundle A imports from B and i_B the number of packages B imports from A. As the dependency between two nodes increases, the resulting unloaded spring length c_2 decreases. The user defined variable c_3 represents the lower bound on the spring length. It should be noted that F_a is applied to both nodes, only if they are interdependent. If i_A or i_B is zero, the attraction force is applied only to one node.

For nonadjacent nodes a repulsion force F_r is introduced.

$$F_r = \frac{c_4}{d^2} \quad (4)$$

The repulsion force is described by an inverse-square law, while its relative strength can be controlled with the user defined variable c_4.

Once all forces for a particle have been accumulated, they are applied to determine the next position of the particle. This is done under the assumption of a constant time step Δt and a particle mass of $m = 1$.

Visualization of Service Dependencies. The main entities of the OSGi service layer are *service interfaces* and *service components*. As these components are linked to Java class types, we visualize them as special building types. To visualize the relationships between the service entities, we introduce the service connection node. These nodes hover above the service interface and service component buildings at a certain height and act as connection points for them. Each node has a visual downward connection to its parent building in order for the user to quickly locate its associated service entity. There are three distinct types of nodes. Service interface (*SIN*), service provide (*SPN*) and service reference nodes (*SRN*). They are assigned to different *service slices*, where each slice resides at a specific height.

SIN nodes assume a central role as they form connections to the other two node types. All SPNs and SRNs connected to a SIN form a *service group* and are members of the same service slice. Only a few service groups are assigned per service slice. This reduces the visual complexity, as the nodes and their connections are evenly distributed over the available height dimension. Due to this design, there are no connections going across individual height layers. Connection crossing can only occur between the service groups that reside in the same service slice. However even this can be reduced as the individual service groups are independent and can be assigned to arbitrary slices.

3.3 Virtual Table

We choose a virtual table metaphor to integrate the software visualization into the virtual environment. The visualization is presented on top of a virtual table situated in an arbitrary room. The entire content of the visualization is confined to the extents of the table. In contrast to a real-world scale visualization, which is more likely to cause a feeling of presence of being inside the data, the table metaphor allows a more strategic/analytic view of the data. Although the table size may vary based on user preference, the metaphor itself imposes a restriction on the size of the visualization space. However this limitation is not a disadvantage, since it enforces to show the visualization in a space saving representation. While it can be helpful to see the fine grained details of software artifacts, it is the higher abstraction levels which contribute mostly to program and architecture comprehension.

The virtual table metaphor provides a transparent transition between individual abstraction levels, as the user does not experience any relocation, since only the visualization in the confinements of the table has to be changed without altering the virtual room around it. This reduces user disorientation and motion sickness, as the room always provides a stable frame of reference [2,12]. This is especially important for the usability of the system for software comprehension, as users can stay longer immersed in the virtual environment without interrupting their train of thought.

As has been noted in Sect. 3.1, the abstraction levels should be directly connected to the relative scale of the elements, which translates to an up or down scaling of the visualization itself. If, due to a high scaling factor, parts of the visualization extend beyond the confined bounds of the table, they will not be displayed. Only content inside of the table bounds is visible. This poses a problem for the display of fine granular software artifacts while preserving their surrounding context. However it is the trade-off when using this metaphor.

On the other hand, the limited visualization volume does not force the user to move around excessively in the virtual environment to view the desired information. This makes the metaphor also very suitable for a seated or standing VR experience, which can improve user comfort and reduce the dependency on VR hardware capable of precise positional tracking.

3.4 Virtual Environment

Although the entire software visualization is displayed in the compounds of the table, the enclosing room plays an important role. In order to maintain the plausibility of all "magic" interactions the table is capable of, a futuristic design is chosen, where the table is augmented with holographic functionality. With the software visualization interpreted as a hologram, the room for plausible interactions is very large. This functionality is implemented by simply discarding all rendered fragments of the software visualization, that exceed the table radius. When designing the environment, we avoided introducing an excessive brightness contrast. This helps in minimizing the "godray" effect, attributed to the used Fresnel lenses.

4 Interaction and Navigation

To enable the user to fully focus on software comprehension, the cognitive load introduced by navigating and interacting with the virtual environment must be minimal. This requires both activities to be intuitive and natural. We build upon the available positional information of the input devices and integrate all interaction possibilities into the environment itself. This reduces the reliance on various button presses and keeps the user interface simple (Fig. 5).

The software visualization appears in the confines of a virtual table, which is placed inside a room. Due to the use of virtual reality and its inherent navigational advantages, the user can walk around the table and inspect the visualization from different perspectives. However this navigational freedom has its limits when inspecting elements up close, as the human visual system has a limit to the distance it can focus on and fuse a stereoscopic image. Therefore it is crucial to be able to additionally manipulate the visualization itself.

The displayed island system has great resemblance to a cartographic map. Thus the proposed manipulation scheme should be familiar to the user, from the usage of digital maps. Our navigational technique encompasses translation, rotation and scaling. It is very similar to the "Two-Handed Interface" technique described by Schultheis et al. [15]. In contrast to their work we constrain the rotation to one axis (Fig. 6 bottom). The scaling operation is especially important, as zooming is directly tied to the transition between the individual abstraction

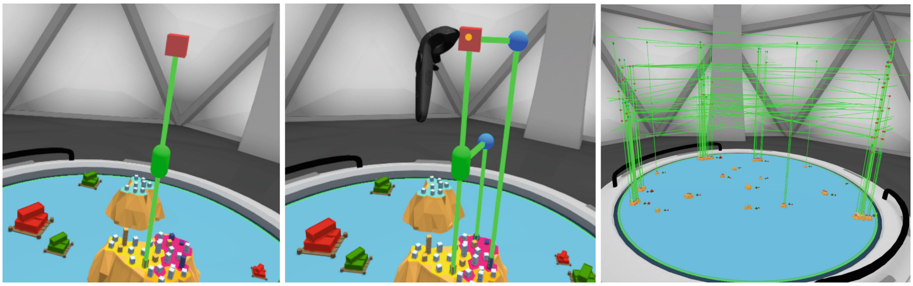


Fig. 5. (left) The two service nodes signalize that the component provides, as well as references a service interface. (middle) The connections to the two service interfaces are shown. Both are placed at different heights as the blue service interface nodes are assigned to two distinct service slices. (right) Multiple service connections shown simultaneously. (Color figure online)

layers of the software architecture. This mode of navigation basically follows a level of detail scheme, where the elements belonging to a specific layer can be interacted with, as soon as they are large enough for the user to see and select.

The visualization can be translated along the axis defined by the table plane. This usually results in left, right, forward, and backward panning, while the translation in the height dimension given by the table normal is prohibited. To apply the translation, the user grabs the visualization and drags it in the direction he wishes to translate, releasing it again when finished.

To perform rotation and scaling, the visualization needs to be grabbed with both controllers. Once grabbed, a virtual pivot point P is established between the controllers. Moving the controllers away from P, along the surface plane of the table, results in a scale increase. Moving them closer towards P decreases the scale. Both actions can be interpreted as a "stretching" or "compressing" of the visualization. In order to rotate the visualization, both controllers are moved in a circular motion around the pivot point. As with a cartographic map, the rotation is constrained to the axis defined by the normal of the table surface. This control scheme allows both scaling and rotation to be performed simultaneously, while P acts as the transformation origin.

Due to the direct manipulation technique, the visualization is more likely to be interpreted as an object and not as part of the stable reference frame [6], reducing discomfort upon navigation.

4.1 Displaying Textual Information

Displaying the names of individual elements is a crucial aspect of software visualization, as it establishes a connection to the underlying software artifact. The display of textual information in a virtual reality environment however, is a challenging task, due to the severe resolution limitations of current head-mounted displays (HMDs). For text to be clearly readable, it has to occupy a

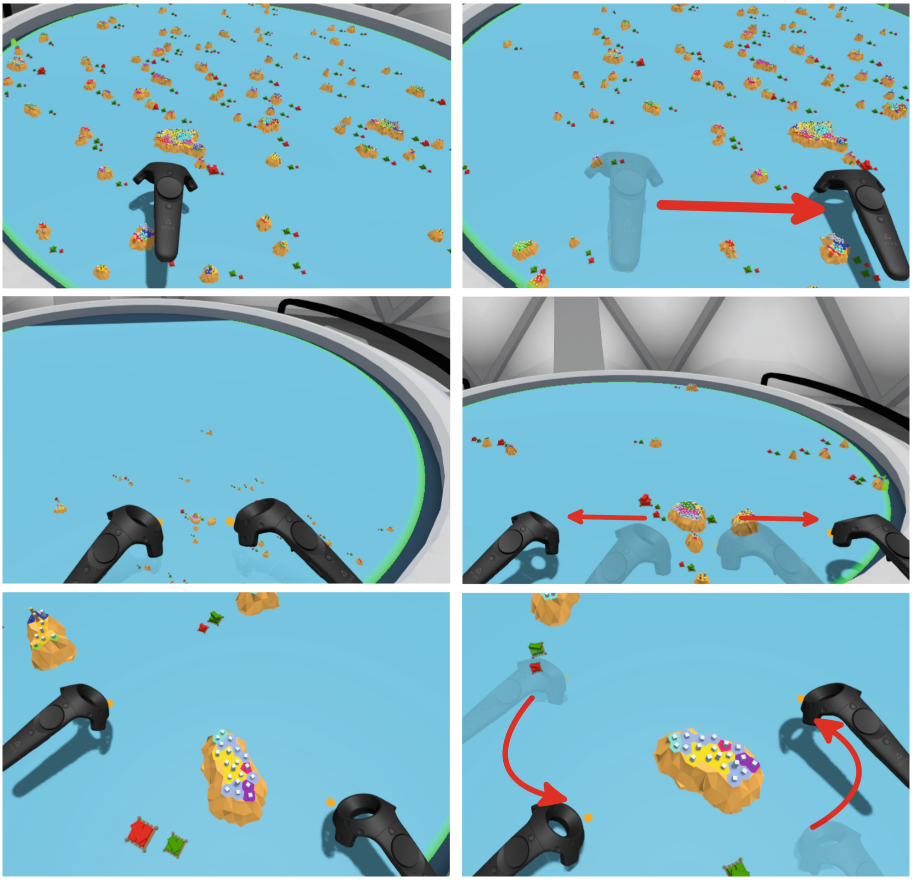


Fig. 6. Navigational operations: (top) Translation. (middle) Scale. (bottom) Rotation.

significantly larger angle in the users field of view, which prohibits the display of large quantities of textual information in the virtual world.

Ideally, the user should be able to know which element he is looking at, without introducing any additional effort. Constantly displaying the text labels of every element however, is not a good solution. The required text size would quickly result in cluttered, overlapping labels, which would increase the overall visual complexity by a large amount. Instead, we display only the text labels of elements which are being hovered over by the controllers. This provides a better control and frees up the HMD for performing only navigational tasks, as opposed to gaze based selection. To display the names of elements further away, a laser pointer functionality is added.

Each time a label is displayed, it adjusts its scale to take up a constant amount of display space, irrespective of its actual distance to the user. Thus, ensuring a consistent readability. However once a label is displayed, it will not further change its scale. This allows the labels to be perceived as 3D objects anchored in the virtual environment.

4.2 Virtual Personal Digital Assistant

While world space anchored text labels are good for displaying object names, they are not suitable for the display of larger amounts of text. However such a functionality is greatly needed, as some information are best presented in their textual form.

A virtual monitor or panel can be anchored somewhere in the environment. When the user interacts with diverse elements, additional information is displayed on this panel. To ensure a good readability, the panel has to be very large and must be placed somewhere, where it is not occluded by the environment or vice versa. To avoid the placement problem, we anchor the display to the virtual body of the user. More specifically, to his hand. This way the panel avoids occlusion problems through the environment, as the user can reposition the panel at any time, without any cognitive effort. The benefit of a large information storage capacity is preserved, as the close proximity of the panel results in large viewing angles.

The panel is attached to the non-dominant hand of the user, so it can be interacted with, by use of the dominant hand. This represents a "double-dexterity" interface, as the interacting hand can be brought to the panel, or the panel to it (or both) [6]. The panel can be thought of as a *virtual Personal Digital Assistant (VPDA)* or tablet (Fig. 7). To avoid unnecessary occlusion and unintentional interactions, the VPDA is disabled per default and has to be explicitly activated by the user. This is done by turning the underside of the controller, or the palm of the hand, towards the user. Inside the VPDA, a classical tabs and windows system is employed, to organize information as well as provide additional functionality.

Fig. 7. To activate the VPDA, the underside of the controller is rotated into the users field of view, providing access to additional textual information and functionality.

5 Implementation

We developed our software prototype, *IslandViz*, using *Unity3D*. The targeted HMD is the *HTC Vive*. We validated our approach by visualizing the OSGi-based software project `RCE`[1]. To extract all relevant information from the software we used a tool based on the work of Seider et al. [16].

6 Related Work

Maletic et al. [8] presented a visualization of C++ code in a virtual environment. Classes are represented as floating platforms upon which additional geometric shapes are placed to visualize attributes and methods. While inheritance is implemented via platform adjacency, other dependency types use explicit connections. The presented system is displayed inside a CAVE environment and showcased only very simple software systems.

Fittkau et al. [4] proposed an approach for a live trace visualizations using a city metaphor in virtual reality. The visualization is presented to the user in a head mounted display (*Oculus DK1*). Since the used hardware does not incorporate any positional tracking, the visualized content is additionally transformed via gesture based controls. In combination with a gaze driven pointer, objects can be selected and interacted with.

Schreiber and Brüggemann [14] introduced an approach for visualizing software modules using the metaphor of electrical components. Modules are represented as blocks and the containing packages are stacked on top of each module. The stacked modules are visualized in virtual reality by various placement algorithms based on the relationship between modules. Modules and packages can be selected an interactively explored by showing service modules, classes, and dependencies between modules and packages.

Recently, Merino et al. [9] and Vincur et al. [20,21] presented a VR visualization for object oriented software (Java, C/C++) using a city metaphor. The approaches rely on VR hardware capable of positional tracking, as the main navigational mechanism is physical movement and interaction is based on the controller positions. In contrast, the main navigational mechanism in our work is the explicit transformation of the visualization itself and is therefore independent of the available physical tracking space. Additionally, we also support positional tracking.

7 Conclusion and Future Work

We presented our approach for exploring OSGi-based software architectures in virtual reality. Based on user feedback, we conclude that a software visualization in VR has great potential in the educational field, as insights into the world of software development can be, almost casually, conveyed to the public.

---
[1] http://rcenvironment.de/.

Further evaluation is needed to determine the practicability of the approach in aiding software comprehension tasks. For this, additional software metrics and functionality need to be incorporated, to construct more realistic scenarios of use.

An advantage of the islands metaphor is the presence of the ocean as the "base plane", spanning over the individual elements. Water possesses interesting optical properties, which could be used for filtering tasks. This could allow the user to reduce the visual complexity by submerging specific islands under the ocean.

A hand based interaction scheme should be investigated and compared to the existing controller scheme. Additionally, a hand based interaction would allow the use of a real physical table prop. This prop would be aligned to the virtual table and provide a form of passive haptic feedback, which has the potential of increasing the users presence in the virtual environment.

Due to the choice of the table metaphor, the visualization should be, at least on a conceptual level, easily portable to an Augmented Reality medium. Its performance and usability in it, could be the topic of future work.

With the recent release of JDK 9, Java finally receives native support for modules. With a slight change, *IslandViz* will be able to visualize a large quantity of future Java based projects, which is an exciting prospect.

References

1. Caserta, P., Zendra, O.: Visualization of the static aspects of software: a survey. IEEE Trans. Vis. Comput. Graph. **17**(7), 913–933 (2011). https://doi.org/10.1109/TVCG.2010.110
2. Duh, H.B.L., Parker, D.E., Furness, T.A.: An "independent visual background" reduced balance disturbance envoked by visual scene motion: implication for alleviating simulator sickness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 85–89. ACM (2001)
3. Eades, P.: A heuristic for graph drawing. Congressus Numerantium **42**, 149–160 (1984)
4. Fittkau, F., Krause, A., Hasselbring, W.: Exploring software cities in virtual reality. In: 2015 IEEE 3rd Working Conference on Software Visualization (VISSOFT), pp. 130–134, September 2015
5. Irani, P., Ware, C.: Diagramming information structures using 3D perceptual primitives. ACM Trans. Comput.-Hum. Interact. **10**(1), 1–19 (2003). https://doi.org/10.1145/606658.606659
6. Jerald, J.: The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool, New York (2016)
7. Koschke, R.: Software visualization in software maintenance, reverse engineering, and re-engineering: a research survey. J. Softw. Maintenance **15**(2), 87–109 (2003). https://doi.org/10.1002/smr.270
8. Maletic, J.I., Leigh, J., Marcus, A., Dunlap, G.: Visualizing object-oriented software in virtual reality. In: Proceedings 9th International Workshop on Program Comprehension, IWPC 2001, pp. 26–35. IEEE Computer Society, Washington, DC (2001)

9. Merino, L., Ghafari, M., Anslow, C., Nierstrasz, O.: CityVR: gameful software visualization. In: 2017 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 633–637. IEEE (2017). http://scg.unibe.ch/archive/papers/Meri17c.pdf
10. Mili, R., Steiner, R.: Software engineering - introduction. In: Revised Lectures on Software Visualization, International Seminar, pp. 129–137. Springer, London (2002). http://dl.acm.org/citation.cfm?id=647382.724792
11. Parks, M.: American Flow Mapping: A Survey of the Flow Maps Found in Twentieth Century Geography Textbooks, Including a Classification of the Various Flow Map Designs. Georgia State University (1987). https://books.google.de/books?id=mgRENwAACAAJ
12. Prothero, J.D., Draper, M.H., Parker, D.E., Wells, M.J.: The use of an independent visual background to reduce simulator side-effects. Aviat. Space Environ. Med. **70**, 277–83 (1999)
13. Robertson, G.G., Card, S.K., Mackinlay, J.D.: Information visualization using 3D interactive animation. Commun. ACM **36**(4), 57–71 (1993). https://doi.org/10.1145/255950.153577
14. Schreiber, A., Brüggemann, M.: Interactive visualization of software components with virtual reality headsets. In: 2017 IEEE Working Conference on Software Visualization (VISSOFT) (2017)
15. Schultheis, U., Jerald, J., Toledo, F., Yoganandan, A., Mlyniec, P.: Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3D tasks. In: 2012 IEEE Symposium on 3D User Interfaces (3DUI), pp. 117–124, March 2012
16. Seider, D., Schreiber, A., Marquardt, T., Brüggemann, M.: Visualizing modules and dependencies of OSGI-based applications. In: 2016 IEEE Working Conference on Software Visualization (VISSOFT), pp. 96–100. IEEE (2016)
17. Storey, M.A.D., Wong, K., Muller, H.A.: How do program understanding tools affect how programmers understand programs? In: Proceedings of the Fourth Working Conference on Reverse Engineering, pp. 12–21, October 1997
18. Tavares, A.L., Valente, M.T.: A gentle introduction to OSGI. SIGSOFT Softw. Eng. Notes **33**(5), 8:1–8:5 (2008). https://doi.org/10.1145/1402521.1402526
19. Tobler, W.: Experiments in migration mapping by computer. Am. Cartographer **14**, 155–163 (1987)
20. Vincur, J., Navrat, P., Polasek, I.: VR city: software analysis in virtual reality environment. In: 2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 509–516, July 2017
21. Vincur, J., Polasek, I., Navrat, P.: Searching and exploring software repositories in virtual reality. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, VRST 2017, pp. 75:1–75:2. ACM, New York (2017). https://doi.org/10.1145/3139131.3141209
22. Ware, C., Franck, G.: Evaluating stereo and motion cues for visualizing information nets in three dimensions. ACM Trans. Graph. **15**(2), 121–140 (1996). https://doi.org/10.1145/234972.234975
23. Yang, M., Biuk-Aghai, R.P.: Enhanced hexagon-tiling algorithm for map-like information visualisation. In: Proceedings of the 8th International Symposium on Visual Information Communication and Interaction, VINCI 2015, pp. 137–142. ACM, New York (2015). https://doi.org/10.1145/2801040.2801056

Interaction in Virtual Environments - How to Control the Environment by Using VR-Glasses in the Most Immersive Way

Barbara Streppel[1](✉), Dorothea Pantförder[2], and Birgit Vogel-Heuser[2]

[1] MaibornWolff GmbH, Munich, Germany
barbara.streppel@maibornwolff.de

[2] Institute of Automation and Information Systems, Technical University of Munich, Munich, Germany

Abstract. Not only in the gaming industry is Virtual Reality (VR) the new way to give users a new experience – in engineering or production plant operation we also see first attempts at finding innovative ways of visualizing data or training plant staff. This is necessary because processes are getting more and more complex thanks to higher interconnection and flexibility. This paper presents actual possibilities of interacting with a virtual environment and provides three concepts for immersive interaction. We also show the results of an evaluation of these concepts at the end of the paper.

Keywords: Virtual reality · Interaction · Virtual environment

1 Introduction

The world is getting more and more complex and needs new ways to display and further understand this complexity. First approaches for using Virtual Reality in this context exists. The advantage of representing complex problems in VR is the increased sense of presence [1]. Presence describes the feeling of being physically present in a non-physical virtual environment (VE) and is separated into immersion and involvement.

Immersion is a crucial factor for presence and can be influenced by interaction with the VE. Previous works indicate that operators in a monitoring task spotted more problems if they were able to interact with a 3D visualization [2].

Involvement is mainly influenced by the environment of a system – e.g. desktop PC, VRWall/Cave or VR-Glasses [3]. Hence, using VR to visualize data, processes or software structure supports the plant staff and software developer in analyzing and understanding of such complex problems. Especially the ability to look around freely and without using a mouse increase the perception of the environment [4].

An important part of every human-operated application is the human-machine-interface. How to design and visualize a 2D-interface has been extensively explored and there are existing guidelines for this, e.g. in ISO 9241. However, every developer/UX-designer must decide on their own how to design and visualize data and interfaces in real 3D/VR.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 183–201, 2018.
https://doi.org/10.1007/978-3-319-91581-4_14

The first part of this paper summarizes the possibilities of how to interact with a virtual environment (Sect. 2). Following this we verify if we can transfer requirements of a 2D-application into VR for an intuitive and natural interaction (Sect. 3). Based on these requirements we divide different interaction methods into explicit interaction concepts (Sect. 4). In Sect. 5, we explain the basics of our software prototype, followed by Sect. 6 where the structure, provision and analysis of the evaluation is be reflected. A summary, conclusions and an outlook on further research potentials is given in Sect. 7.

2 Related Work

There is much research on Human-Machine-Interfaces or Human-Computer-Interactions for 2D applications, which we present in this section. Existing works on interaction concepts in virtual reality are also presented here. At the end of this chapter, we evaluate existing controls and try to divide them into subcategories.

2.1 Human-Machine-Interface

Every task a machine should execute in response to human input needs an interface for transmitting this information. This can be a simple, physical button or a software user interface. Human Machine Interfaces (HMI) include all points of contact between machine and human. This contains everything that a machine displays to the user and – vice versa – all user inputs for the machine.

The research field of ergonomics follows up with a good handling for HMIs. For this work we take a closer look at the field of software ergonomics. The most important rules and recommendations for a good usage of an 2D user interface are described in ISO 9241. The norm defines a guideline for human computer interactions, e.g. amongst others guidance on dialogue principles or guidance on usability [5].

To rate the usability of a software you can use the questionnaire laid out in ISO 9241/110, which uses the seven ergonomic principles, which are: suitability for the task, suitability for learning, suitability for individualisation, conformity with user expectations, self-descriptiveness, controllability and error tolerance. Nielson (1994) recognized similar factors for the quality of a human machine interface: visibility of system status, match between system and real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, help users recognize, diagnose and recover from errors, help and documentation [6].

In this paper we try to figure out which of these 2D-principles can be transferred to a 3D world and can be used to create an intuitive and natural user interaction concept for VR (Sect. 3.2).

Even if a software or UX developer takes all these principles into account, he needs to perform user tests to get the final interface design. This is because every application needs a special interaction concept, which depending on its working task. To create a working interaction concept we performed some user tests to receive feedback on our concept ideas before implementing the final prototype for the final evaluation.

2.2 Virtual Reality – Interaction

Interactions with a virtual environment are not the same as with a real environment. Scenarios which partially allows more than in reality and the missing haptic feedback in a virtual world require new concepts of interaction.

In some cases, voice control is used to control a VE. More common is the usage of controller or gesture control. Generally, there are at least two steps for the user to interact with objects in a VE [7]:

1. *Selection*
   a. Directly grabbing the object (e.g. by pressing a button) or
   b. Selecting the specified object (e.g. by ray cast) and then grabbing it (e.g. by pressing a button)
2. *Manipulation*
   a. Changing the position of the object
   b. Changing the orientation of the object

Mine defined two more forms of interaction: *movement* and *scaling*, and from these derived a fifth form, *virtual menu and widget* [8].

Several papers provide various classifications for moving. Ware gives a general classification of moving in a 3D world [9]. He divides moving into *walking* and *flying*, whereby moving has two degrees of freedom and flying has three. Mine introduced several other principles, one is called *dynamic scaling*. Here, the user can minimize the world, so he has to move a minimal way to reach his goal [8].

Sasse et al. use a division into two concepts called *world-in-hand* and *eyeball-in-hand*. With *world-in-hand* the user is fixed in position and can move the world around him. With *eyeball-in-hand* the world is fixed, and the user moves around the world [10]. This division can also be used for *scaling* and *manipulation*. In an application the movement method should be consistent for the user, i.e. either he moves or the world does.

Furthermore, Mine introduced three categories of how to implement the existing forms of interaction:

1. *Direct User Interaction*: Hand tracking and gesture recognition to map real user action to a resulting action in the virtual world.
2. *Physical Controls*: Real physical objects to interact with the virtual world, like a steering wheel in a driving simulation.
3. *Virtual Controls*: Interact with the environment by using Virtual Control elements.

Direct User Interaction could also be named gesture control. There is no agreed definition of the word gesture. Kurtenbach und Hulteen described a gesture as the motion of a part of the human body that contains information. For them pushing a button isn't a gesture, because it doesn't contain information [11]. For Gartner the most important part of gesture control is that the user doesn't touch the physical world [12]. Furthermore, Harling and Edwards consider only gestures of hands without the interaction with a physical interface. Therefore, they differentiate between hand poses and positions. Both can be static or dynamic. This leads to the following classification in Table 1 [13]:

Table 1. Classification of gesture of Harling and Edwards

Hand pose	Hand position	Example
Static	Static	Peace-Sign
Static	Dynamic	Waving goodbye
Dynamic	Static	Grabbing
Dynamic	Dynamic	Throwing

2.3 Existing Controls

There are several VR applications, most of them games. To get an overview of existing controls we investigated other VR applications. Selecting and manipulating objects are most of the times implemented in the same way. Only moving is different in many applications.

Apart from these classifications, most applications use a teleport system for moving the user, thereby avoiding motion sickness. For this the user can use one button to create a laser pointer and release it, or use a second button to teleport to the appearance point of the laser pointer. Sometimes the user can define his direction in which he wants to look after teleporting [14]. Another possibility is pausing the application at the end of the room tracking and let the user rearrange his avatar. This can be done by turning the head of the user, while the avatar remains stationary. In the game Runes the user can switch between first and third person view [15]. To avoid motion sickness, the user stands at a fixed point when he is using third person and moves his avatar with a remote control.

For grabbing things, the principles of Sect. 2.2 are used. Most of the time the user can press a button and hold it down or click the button to pick up a subject and click again to release it. Google Earth uses the first form for letting the user move over the world. This is like grabbing and pulling a fixed object.

Interacting with UI elements, like in a menu, the investigated applications use two different concepts. The user can use a laser pointer to select a UI element or he touches the element directly with the controller, which commonly gives haptic feedback. To use the selected element, he has to push a controller button.

Some controls are made for gesture control, so they can be used with a camera for hand or gesture recognition. There the user for example can grab subjects by using a fist or a pinch gesture.

In some cases moving with gestures is the same as with a controller, where the user grabs the world and pulls himself around. Another possibility is to have points of interest floating around the user, which can be selected. Furthermore, a neutral hand position can be defined and the distance between the hand and the camera defines the velocity of moving. Hence the user can move forward and backward [16, 17].

3 Requirements Analysis

The goal of this paper is to evaluate which are good interaction concepts for a VE. A good concept should be intuitive, natural and easy to use. To achieve this we set up requirements for interaction in VR. To be able to evaluate the different concepts there should be a VR scenario in which the evaluation participants can test them. This scenario was chosen to be a so-called “software city”[1]. After explaining the initial situation we will describe the derived requirements.

3.1 Application Example

We decided to use a software city as an application example for the different concepts. An example of a software city can be seen in Fig. 1. A software city is a representation of software code as a city. The streets of the city constitute the folders of the software code, and buildings constitute the classes. Height, base area and building color can represent different metrics, such as lines of code, complexity or test coverage. The user can therefore directly see where a large class is or where there are not enough tests [18].

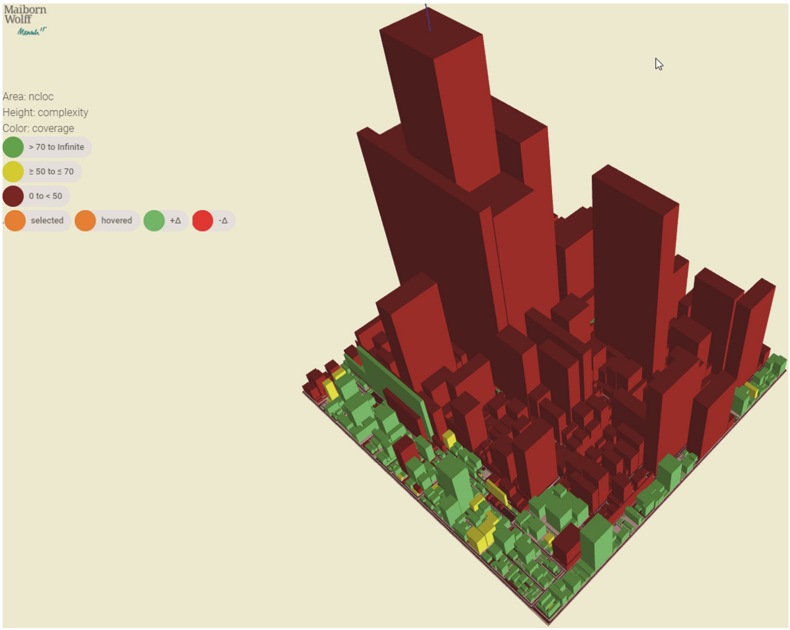


Fig. 1. SoftwareCity desktop version

A desktop application enables the transformation of code into a software city. The user can move, rotate and scale the city. Furthermore, he is able to select a building and get more information about the related class.

The software city is used to analyze software code. Therefore, a software developer can see the whole of his software and the relations between different classes as well as explaining the complexity or quality of the software e.g. to a manager who has no deeper understanding of software development.

[1] Now CodeCharta.

3.2 Requirements

As the VR version should have the same interaction possibilities as the desktop version, the following should be implemented:

- Movement between city and user
- Rotation between city and user
- Scaling between city and user
- Selecting a building to get more information
- Interacting with the UI elements which contain the information

The interactions should be immersive, so the VR-system needs hand tracking or a handheld controller. Because moving is a requirement for the application, motion sickness should be minimized. The VR-system should be portable so it can be taken to customers.

To get the advantages of VR, such as an increased perception of the environment, the application must have a high immersion. This is achieved through a good display and the application's interaction fidelity [19]. Display fidelity is affected by the VR-system used and its accompanying computer system. To have a high level of interaction fidelity, the interactions should be natural and intuitive, so that they are as near as possible to real or other known interactions.

Furthermore, all interactions should follow the rules of dialogue principles, as defined in ISO 9241-110. These rules are for desktop applications, but can be adapted for VR interaction controls:

- Suitability for the task: the control supports the user by executing his task
- Suitability for learning: the control offers instructions
- Suitability for individualization: the control can be customized
- Conformity with user expectations: the control is consistent and follows known conventions
- Self-descriptiveness: the user knows at every point what he is doing and can do
- Controllability: the user can go back and can undo actions
- Error tolerance: the control is robust against invalid user input and tries to avoid it.

Each of these rules can be adapted for VR interaction controls. It's unclear how necessary different rules have to be considered for the interactions to be implemented. E.g. for the rule of controllability the user can undo unwished interaction, e.g. a too far scaling, by scaling it again in the opposite direction and doesn't need a back button.

4 Concept

After defining the requirements for the application, we decided to define three different interaction concepts. Thereby, the classification of the concepts is inspired by the classification of Mine (Sect. 2.2). Hence, there are a gesture control (direct user interaction), a control by physical controller (physical controls) and a control by virtual UI elements (virtual controls). It should be noted that we considered voice control but didn't implement it as scaling or moving a city by voice felt unnatural. For following researchs it would be interesting to implement parts of a VR application as voice control.

Because of the portability requirement the concepts should be implemented for a head mounted display (HMD). At the beginning of the work only the HTC-Vive had tracked controllers, so we used it as HMD [20]. To enable gesture control we used Leap Motion [21].

Below the procedure of creating the different concepts is presented and the usage of specific interactions is explained. For creating a natural and intuitive interaction concept we perform a preliminary experiment with some participants.

4.1 Preliminary Experiment Defining Suitable Interaction Elements

To have the controls as a product of human imagination and therefore natural and intuitive for most people, we created a previous version of the prototype. With it we conducted interviews with six participants. The participants had different knowledge about VR, the software city and controller usage. Features implemented in the pre-prototype were scaling, rotating and moving the city with controller and UI elements.

Every participant was interviewed alone and by using a guideline with pre-defined questions. By using these questions, the participants were lead through the interview. They tried out the pre-prototype and with this we discussed the points of discussion shown in Sect. 4.2.

4.2 Discussion of Different Interaction Elements

As described in Sect. 3, the user should be able to move around in the city. The user could do this by moving himself, or he could move the city around him as in the principles of Sect. 2.2 world-in-hand or eyeballs-in-hand. By using the HTC-Vive with its room scale tracking the user can also walk by himself in the provided space.

The same principles are important for rotating. Because of the tracking of the user, he could just turn his head to see the city from another perspective. So, perhaps he doesn't need the possibility to rotate it. Generally it's important for manipulation to know the center of rotation and scaling, i.e. if the city rotates/scales around the user or around its own center.

There were three topics of discussion before starting the implementation of the real prototype. Each topic has various implementation possibilities:

1. Moving:
   a. The user moves inside the city
      (1) In real world by room scale tracking
      (2) By using moving elements
   b. The user moves the city around him
2. Rotation:
   a. The user doesn't need rotational elements, because he can turn his head
   b. The user can rotate the city

    (1) Around himself
    (2) Around the center of the city
3. Scaling:
    a. Scaling is around the user himself
    b. Scaling is around the center of the city.

To determine which possibilities to implement, we interviewed the six participants, as described in Sect. 4.1. At this point there were a few interactions implemented, so they could try out some possibilities. Users consistently noted that the movement of the city relative to the user was crucial. There are two different main positions for the user: (A) The user is outside the city and the city is small right before him and (B) he stands inside a large scaled city. Table 2 shows the decisions of the participants which interactions should be possible for which position.

Table 2. Possible interactions depending on the users position

	Position A	Position B
Moving	1a & 1b	1a & 1b
Rotating	2b(2)	2b(1)
Scaling	3a	3a

Generally, the suggestions of the participants point to a manipulation of the city not of the user. There should be a way to move/scale/rotate the city, because the user can move and rotate himself without controls just by real moving or turning his head. However, moving the city and himself in some way should be possible, because there will be not enough space to walk around in a large city inside a small testing room.

After trying the first parts of the prototype they also mentioned that the cables of the HTC-Vive were disturbing when turning their head, and it would be better to always have the ability to rotate the city. If the user has the possibility to move the city it would be enough to rotate the city around the user, so he doesn't lose track of his position inside it.

However, for rotating there should be two possible rotation centers. If they are in position B, the city should rotate around the user. In position A, it should rotate around the city itself, so the city can be viewed from all sides.

During the interviews we asked the participants also which gestures they would prefer to interact with the software city for the required interactions.

4.3 Direct User Interaction

In this concept all interactions are made with gestures. In this case gestures are a combination of the pose of the hand and its movement. Both will be recognized by a leap motion camera [21]. The leap motion can recognize the position of a hand and its fingers, as well as the direction the palm and the fingertips are pointing in. It follows that the gestures have to be unique and easily recognizable for a leap motion camera.

Gestures should be as natural as possible so the user can remember them. Don Norman, Director of the Design Lab of The University of California and usability specialist, considers just a handful of gestures to be natural for him. The rest have to be learned and remembered later. This is only possible for far fewer gestures than would be required [22].

Moving. Most of the interviewed participants wanted to move the city and not themselves. Only one suggested a gesture control for moving himself inside a fixed city. We conclude that in this concept it's enough to only implement a way of moving the city around the user. The most commonly suggested way to do this was to grab the city and drag it around. For grabbing, the hand has to be clenched to a fist.

Rotating and Scaling. For rotating and scaling we pursued two strategies: either using one hand or two. To rotate or scale the city around its own center, the user can grab it with both hands, move them relative to each other for rotating, and pull them apart for scaling. This is inspired by the way a map is manipulated on a touch device, e.g. a smartphone, where the user moves two fingers for rotating und scaling a map.

To rotate the city around the user himself, he has to open one hand and point with his palm in the direction he wants to rotate the city (Fig. 2 left). For scaling he has to point his palm into in the direction of the camera. If he points away from the camera and moves his hands, the city is minimized. The other way around maximizes the city (Fig. 2 right).

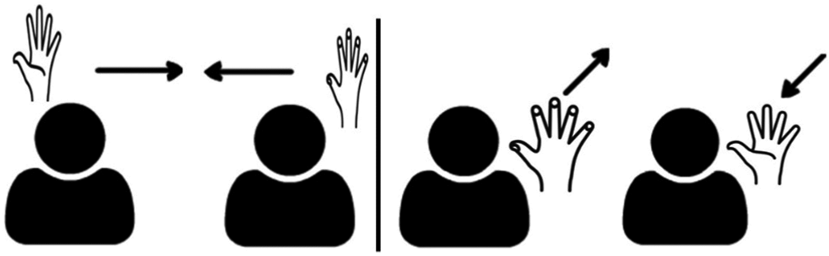


Fig. 2. One-handed gestures to rotate (left) and scale (right) the city around the user

Selecting a Building. To select a building the participants suggested simply touching the chosen building with an outstretched index finger.

Operating the UI. The control of the UI is built up as on a touch device, e.g. a smartphone. For this, the user can use his whole hand in every hand holding position.

4.4 Physical Control

In this concept the user interacts with the virtual environment only by using the controllers of the HTC-Vive. To make it as "physical" as possible, the interactions should be executed, as far as it seems reasonable, only by the buttons of the controller, not by describing gestures with it.

To better understand the different buttons used, Fig. 3 describes the key assignment of the used HTC-Vive Controller. The trigger (Fig. 3 number 7) is used for turning on a laser pointer. For this the user has to slightly touch it. By pushing it up to the end, this means clicking it, the user can select or confirm something. The grab buttons (Fig. 3 number 8) are for grabbing things and the trackpad (Fig. 3 number 2) is used for scaling and rotating the city.

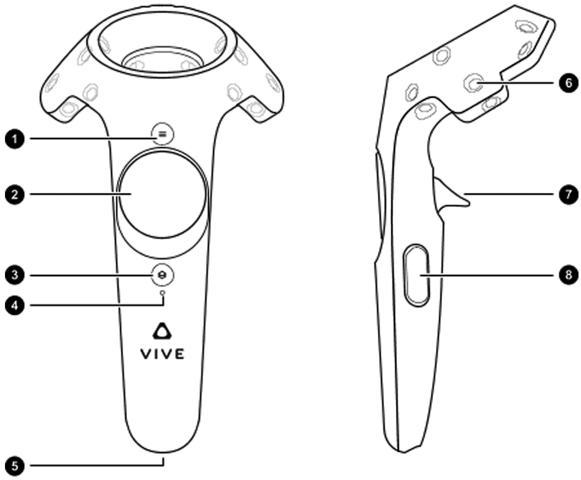


Fig. 3. Key assignment of the HTC-Vive Controller

Moving. Even in this concept, the participants decided that the possibility to move the city, not themselves, is enough. Here, they can use the same method as in gesture control, so they grab the city with the controller and pull it around. Grabbing is executed by pressing the grab buttons of the controller.

Furthermore, the user has the possibility to use a laser pointer to teleport himself through the city. The laser pointer is turned on by touching on the trigger. By pushing the trigger up to the end, this means clicking the trigger, the user conforms that he wants to teleport to the endpoint of the laser pointer.

Rotating and Scaling. For rotating and scaling the trackpad is used. Pushing the upper or lower half of it maximizes or minimizes the city respectively. When the left or right half is pushed, the city rotates. In this context the size of the city is important. The physical control concept constantly recognizes how large the city is. If the city is big, it rotates around the user, and if the city is small, it rotates around its own center. Interview participants said it's enough to scale the city around the user himself if they have the possibility to move the city.

Selecting a Building. To have a self-contained concept, the user here also uses the laser pointer. By touching the trigger, the laser pointer appears. If the user wants to select a building, he can touch it with the beam (Fig. 4) and open the additional information window by clicking the trigger.

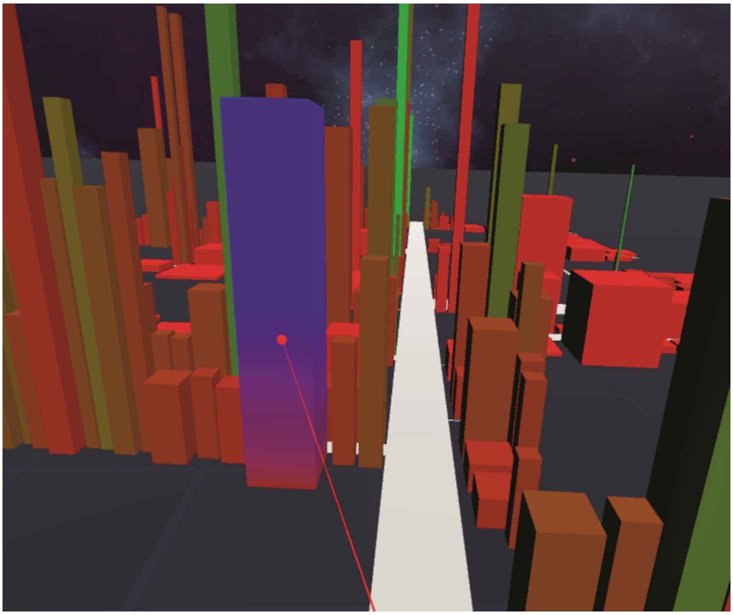


Fig. 4. Selecting a building with the laser pointer

Operating the UI. Also for operating the UI the laser pointer is used. It can be used like a mouse on a desktop PC.

4.5 Virtual Control

In virtual control the user should use virtual elements to manipulate the city. To operate the virtual elements the user should use directly his own hands. It also should be possible to move the elements, so that they do not disturb the user's field of view. The position of the elements can be manipulated by using the trackpad of one of the HTC-Vive controllers. To get moving elements into the foreground and back the user swipes the trackpad from side to side. If he wants to use the scale/rotate elements he swipes from bottom to top or other way around, depending on whether the elements should be in his upper or lower field of view.

Moving. For moving there is a mini map of the city, where the user can see himself and the city from above (Fig. 5 left). To move the city, he can grab inside the map, by clenching his hand to a fist, and pull it in the direction he wants the city to move. This pulling is executed in the coordinate axes of the map, not of the city.

Rotating and Scaling. On the right side of Fig. 5 the elements for rotating and scaling are presented. It was important that these elements are positioned and the user can use them, as in normal touch interfaces. Standard UI elements, e.g. buttons, sliders and knobs, are used for this. This special VR-UI-elements are three-dimensional and can be

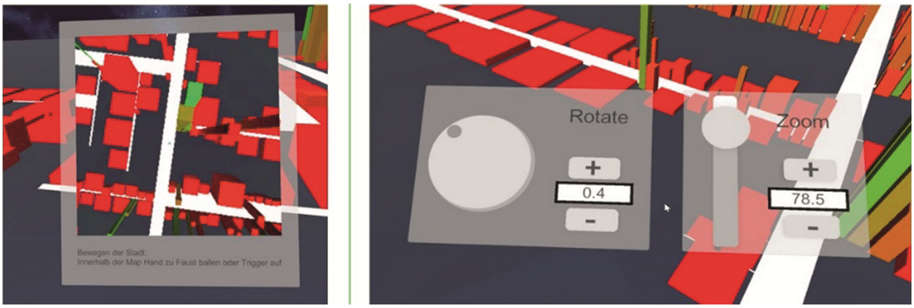


Fig. 5. Moving, rotating and zooming elements (from left to right)

used by pushing them down and then manipulating them. Instead of the missing haptic feedback, there is visual feedback, so the elements change color as soon as they're pushed down. For fast rotating or scaling the user can handle the knob or slider. For fine adjustment he can use the plus and minus buttons.

Selecting a building and operating the UI in this concept works like in the gesture control. Selecting a building can be done by touching it with an outstretched index finger. UI can be operated like a normal touch interface.

5 Implementation of the Software Prototype

Different interaction concepts must be evaluated with a software prototype. The prototype was implemented with Unity3d [23] for an HTC-Vive and the Leap Motion. In Unity we used the SteamVR assets for connecting to the Vive [24]. For connecting the Leap Motion camera with Unity, we used their suggested plug-in. In addition, Leap Motion has prefabricated UI objects we can use for the virtual control concept. We used VRTK[2], a VR toolkit for rapidly building VR solutions in Unity3d, to use the controller more easily, e.g. when creating a laser pointer for teleporting or controlling UI elements [25].

The general structure of the implementation can be seen in Fig. 6. In Unity an implementation is divided into sub modules called "scene". For each concept a different application is built by using a different start scene, that reference to the main scene, the scene with the model of the city and the associated control scene. The *WelcomeScenes* are the entry points for the application. They welcome the user and have a confirmation button. Is this button clicked the *ScaleRotateExampleScene* is loaded where the model of the city is stored. Together with the *WelcomeScene* the associated *PlayerScene* is loaded. This scene contains all data for the different interaction concepts, so that pushing of the button is directly done within the interaction concept. The *MainScene*, which is also loaded when starting the *WelcomeScene*, contains all elements that should be available for every concept and in every scene.

[2] SteamVR Unity Toolkit.

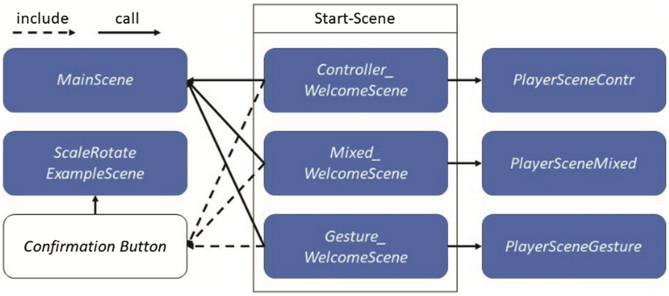


Fig. 6. Procedure organization of the scene flow

6 Evaluation

After an introduction in the structure of the evaluation, there will be an overview over the results (Sect. 6). In the end we conduct an evaluation with 30 participants of different ages and knowledge about VR. The distribution can be seen in Table 3.

Table 3. Distribution of the participants

Alter	Female	Male
18–25	1	5
26–35	4	16
36–45	0	4

6.1 Structure

To give all participants the same information, they get a written and spoken introduction in the beginning. The user is told how the controller works and what the evaluating issue was. The task was to explore a given software city and identify which classes, and therefore buildings, are interesting and why. For this the participants should use all necessary interaction possibilities.

At the beginning of every concept test they get an introduction to the different interaction possibilities (Fig. 7 bottom), which appear in front of the user. When the user performs the given interaction, the explanation disappears and the next information text shows up. Furthermore, when the user made all possible interactions, some informational icons appear on the controller – if any are used – to support the user (Fig. 7 top).

Fig. 7. Help system: top - help signs on the controller, bottom - explanations of an interaction possibility with video (Translation of the text: "Rotate & Scale (1): Make a fist with both hands and grab the city -> Interaction like with Google Maps used with touchscreen".)

At the end of every concept the participants complete a questionnaire based on ISO 9241-110. After testing all three concepts they fill out a questionnaire designed by ourselves for exactly this test case. Questions included "What was your preferred concept and why?" or "What concept would you use in a customer project". Here, also previous experiences with VR, 3D applications or controller games were questioned.

6.2 Results

For better understanding please refer to Table 4 to see how many participants are meant with which description.

Table 4. Information to the description of the number of participants

Description	Number of participants
A few	2–5
Some	5–10
Various	10–20
Many	20–25
The most/almost all	25–30

Observation During Test Execution. There were different results, which we noticed directly while observing the user testing the concepts. For example, various participants had problems interacting with the UI elements. Initially we thought that the user would interact easily with them, because they knew the elements from everyday touch devices, like their smartphone. But the lack of haptic feedback makes interaction more difficult for the user.

Although we discussed the scaling of the city with participants in the first evaluation, the participants in the final evaluation had problems with it. For some the city scaled in the wrong direction. While every interaction in the concepts is based on manipulating the city (except the teleportation in the physical control concept), these participants expect by pushing the trackpad up, that they move up, not the city, and hence, the city gets smaller not bigger.

The gesture control was difficult to learn for most of the participants. They said it was hard to avoid making unintended gestures. They often moved the city when they closed their hands to select a building or while opening them to release the city. The direction of the palm for scaling the city was also confusing for many of the participants. They expected, that each time they move their hand relative to the ground the same happens. However, Leap Motion expects the palm to be in the direction of the camera view. Some users complained about the cognitive effort in this type of gesture control, because of the big focus on hand poses and positions.

In virtual control concept the mini map obtained little approval. Some mentioned that they liked the map, but only for an overview, not for manipulation. They could indicate specific points of interest and move there e.g. by directly touching the map. In particular, the movement in different coordinate systems was confusing for most of the participants.

Comparison Questionnaire. At the end of all concept tests the participants completed a questionnaire to compare the tested concepts. The result of this questionnaire conveys the same picture as the surveillance of the tests. The virtual control has the fewest votes in questions regarding which interaction possibility the users found the best. This question was a multiple-choice question. The answer distribution can be seen in Table 5.

Table 5. Answer distribution of the question: “What concept for * pleases you the most?” (Original: “Welches Konzept für * hat dir am besten gefallen?”.)

*	Controller laser pointer	Controller button	Virtual control	Gestures one hand	Gestures two hands
Move	15	14	3	4	–
Scale	–	18	5	5	12
Rotate	–	14	2	11	15
Select	25	–	–	–	9

The test users voted for the virtual control as least. The rating of gesture control is often similar to the rating of physical control, with a tendency to the physical control. The most pronounced result here is for selecting a building. Some users explained this after the evaluation verbally: because the task was to analyze the city, they wanted a possibility to select a building from further away.

There were also the questions regarding users’ preferred interaction concept in general, and which realization the participants like the most. For this they could rate them on a scale from one to five, where one is the worst and five is the best possible value.

The virtual control has the worst rates average of the three possibilities (2.7 for concept, 3.0 for the realization). The average for the concept rates for the physical control (4.30) is minimally higher than the average for the gesture control (4.0). However, in realization the physical control is much better (4.2) than the gesture control (3.5).

The tendency that the users like the controller and gesture concept the most, is seen in the answers to the question "Which concept would you use in a real project?"[3]. 14 participants answered with gesture control, the other 16 participants with physical control. The reason why they chose this answer was very similar amongst all. The participants could answer this question with plain text. In sum the answers can be seen in Table 6.

Table 6. Count of specific qualities in the reasons why the choose a control

	Physical control	Direct user interaction
Easy to operate	4	2
Easy to learn	1	2
Intuitive	2	4
Go well with VR	0	4
Less mistakes	3	0
Exact/precise	3	0
Fast	2	0

Those who opt for the physical control justified their decision by stating that this form of control worked best, had least sources of mistakes and therefore, is the easiest and fastest control. One participant described his decision with the words "exact", "precise" and "stable". Another justified their answer with the sentence "Intuitive operation which is fast and easy to learn".

A similar reason, but for the gesture control, was "It is the easiest way to operate [...] and easy to learn". Two participants considered the gesture control as "natural", many others as "intuitive". Four mentioned that they decided for gesture control because it's the most obvious and go best with VR.

In particular the physical control convinced with its speed, accuracy and low error susceptibility. The gesture control has its advantage in being felt as intuitive and consistent with VR. A relation to personal preferences of the participants, like if they used 3D- or VR-applications before, was not found.

The only possible relation could be with the age of the participant. Four of five participants in the age of 18–25 chosed the gesture control in this question. Three of four in the age of 36–45 decided upon the controller concept. In the age of 26–35 the decisions were balanced. If this is a real correlation or just coincidence should be tested in a next step.

[3] Original: "Welches Konzept würdest du am ehesten in einem Kundenprojekt nutzen?".

ISO 9241/10 Questionnaire. We also used a customized ISO 9241/10 questionnaire to get more information about the usability of the different concepts. The physical control got the best rating in most of the questions, followed by the gesture control. Only in the category of error tolerance gesture control was the worst. In the category of self-descriptiveness, it was in some questions adjusted to the same height with the virtual control. The largest difference between the valuation of physical control to the other two concepts was seen in the question of time requirement.

7 Conclusion

The aim of this paper was to find out which interaction concepts are the best for controlling a virtual environment without haptic feedback. For this we looked for concepts that are felt as natural and intuitive as possible. After a research of existing controls and concepts we decided to implement and evaluate three different concepts leaned on the concepts of Mark. R. Mine.

Generally, gesture and physical control are liked equally. Physical control, such as using a controller, has the advantage being able to give haptic feedback, and is therefore more precise and easy to use. Because the user can feel different buttons, he doesn't use them accidentally and can learn their usage fast. The existence of help icons and text, which are located near the buttons, also supports the self-descriptiveness of this interaction concept. The possibility of an explicit tracking of the controllers also make it more precise and accurate.

Gesture control has the advantage that it's perceived as natural and appropriated for VR. Although the physical control is the best rated in nearly every question in the ISO questionnaire, half of the participants would use gesture control in a real project. Like every new technology gestures have to be learned and people have to get accustomed to it. If in the future there are well known gestures and the technology for tracking and recognizing a hand is good, the most immersive way to control a virtual environment could be gesture control. Until then everyone has to decide personally whether to go the safe way and use a controller, or if they want natural interaction and use gesture control.

To combine the best of both controls, it's conceivable to use gesture control with controllers. Here, the movement of the controllers would map the movement of a hand. Pushing different buttons could correspond to different hand poses. Also, the further development of controllers with proximity sensors should be observed. There, the user can use the benefits of a controller, like haptic feedback and the usage of buttons, and additional utilize basic gestures.

The results of this work are subjective impressions of the participants, which are influenced by the type of the controls implementation. A user study with more participants, fully developed interactions and a real user task, like finding a specific class with measuring the expired time, is needed to get objective results. Also, an evaluation for another application, e.g. the virtual operation of a machine, could bring new insights how much a specific task influences the evaluation of different concepts.

References

1. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoperators Virtual Environ. **7**(3), 225–240 (1998)
2. Pantförder, D., Vogel-Heuser, B., Gramß, D., Schweizer, K.: Supporting operators in process controll tasks – benefits of interactive 3-D visualization. IEEE Trans. Hum.-Mach. Syst. **46**(6), 895–907 (2016)
3. Ulewicz, S., Pantförder, D., Vogel-Heuser, B.: Interdisciplinary communication and comprehension in factory automation engineering – a concept for an immersice virtual environment. IFAC-PapersOnLine **49**(19), 227–232 (2016)
4. Pausch, R., Proffitt, D., Williams, G.: Quantifying immersion in virtual reality. In: The Art and Interdisciplinary Programs of SIGGRAPH 1997. Visual Proceedings, pp. 13–18. ACM SIGGRAPH, New York (1997)
5. DIN Deutsched Institut für Normung e.V.: Ergonomie der Mensch-System-Interaktion – Teil 110: Grundsätze der Dialoggestaltung. DIN EN ISO 9241-110 (2008)
6. Nielson, J.: Usability Engineering. Kaufmann, San Diego (1994)
7. Bowman, D.A., Hodges, L.F.: An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: van Dam, A. (ed.) Proceedings of the 1997 Symposium on Interactive 3D Graphics, p. 35-ff. ACM, New York (1997)
8. Mine, M.R.: Virtual Environment Interaction Techniques (1995)
9. Nazemi, K.: Information visualization. Adaptive Semantics Visualization. SCI, vol. 646, pp. 17–68. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30816-6_2
10. Sasse, M.A., Johnson, C., Johnson, C.W.: International federation for information processing: human-computer interaction. In: IFIP TC 13 International Conference on Human-Computer Interaction, INTERACT 1999, vol. 1, Edinburgh (1999)
11. Kurtenbach, G., Hulteen, E.: Gestures in human-computer communication. In: Laurel, B. (ed.) The Art of Human-Computer Interface Design, pp. 309–317. Addison-Wesley Publishing Co., Boston (1990)
12. Gartner: IT Glossar. Gesture Control. http://www.gartner.com/it-glossary/gesture-control/. Accessed 29 Nov 2016
13. Harling, P.A., Edwards, A.D.N.: Hand tension as a gesture segmentation cue. In: Harling, P.A. (ed.) Proceedings of Gesture Workshop Progress in Gestural Interaction 1996, 19th March 1996, pp. 75–88. Springer, London (1997). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.52.9492. University of York
14. Robo Recall. https://www.roadtovr.com/robo-recall-mod-adds-full-locomotion-address-room-scale-niggles/. Accessed 8 May 2018
15. Stormborn Studio – Runes: The Forgotten Path, a VR wizard adventure. http://www.kickstarter.com/projects/1501963972/runes-the-forgotten-path-a-vr-wizard-adventure. Accessed 13 May 2017
16. Chastine, J., Kosoris, N., Skelton, J.: A study of gesture-based first person control. In: Mehdi, Q. (ed.) 18th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGAMES), pp. 79–86. IEEE, Piscataway (2013)
17. Colgan, A.: 5 experiments bleeding edge vr locomotion. http://blog.leapmotion.com/5-experiments-bleeding-edge-vr-locomotion/. Accessed 13 May 2017
18. CodeCharta. http://www.maibornwolff.de/codecharta. Accessed 8 Feb 2018
19. McMahan, R.P., Bowman, D.A., Zielinski, D.J., Brady, R.B.: Evaluation display fidelity and interaction fidelity in a virtual reality game. IEEE Trans. Vis. Comput. Graph. **18**(4), 626–633 (2012)

20. HTC Vive. https://www.vive.com. Accessed 8 Feb 2018
21. Leap Motion. http://www.leapmotion.com. Accessed 8 Feb 2018
22. Norman, D.: Gestural Control: the Good, the Bad, and the Ugly. https://www.linkedin.com/pulse/20140320012035-12181762-gestural-control-the-good-the-bad-and-the-ugly. Accessed 30 Jan 2017
23. Unity: https://unity3d.com/. Accessed 8 Feb 2018
24. Steam VR: store.steampowered.com/steamvr. Accessed 8 Feb 2018
25. VRTK. https://vrtoolkit.readme.io/. Accessed 8 Feb 2018

Sensor Data Fusion Framework to Improve Holographic Object Registration Accuracy for a Shared Augmented Reality Mission Planning Scenario

Simon Su[1(✉)], Vincent Perry[1], Qiang Guan[2], Andrew Durkee[1], Alexis R. Neigel[1], and Sue Kase[1]

[1] US Army Research Lab, Aberdeen Proving Ground, MD 21005, USA
simon.m.su.civ@mail.mil
[2] Kent State University, Kent, OH 44240, USA

Abstract. Accurate 3D holographic object registration for a shared augmented reality application is a challenging proposition with Microsoft HoloLens. We investigated using a sensor data fusion framework which uses both sensor data from an external positional tracking system and the Microsoft HoloLens to reduce augmented reality registration errors. In our setup, positional tracking data from the OptiTrack motion capture system was used to improve the registration of the 3D holographic object for a shared Augmented Reality application running on three Microsoft HoloLens displays. We showed an improved and more accurate 3D holographic object registration in our shared Augmented Reality application compared to the shared augmented reality application using HoloToolkit Sharing Service released by Microsoft. The result of our comparative study of the two applications also showed participants' responses consistent with our initial assessment on the improved registration accuracy using our sensor data fusion framework. Using our sensor data fusion framework, we developed a shared augmented reality application to support a mission planning scenario using multiple holographic displays to illustrate details of the mission.

Keywords: Shared augmented reality · Sensor fusion

1 Introduction

The introduction of the Microsoft HoloLens [1], a commercial off-the-shelf augmented reality device, has allowed researchers at the United States Army Research Lab to explore using augmented reality technology for data visualization. The ability to superimpose data generated from a physics-based modeling and simulation into the actual physical environment is a very effective way of showing the results of the simulation. In addition, it is more than likely that a data visualization session is a collaboration between a group of researchers and stakeholders of the project. Shared augmented reality capability would be needed to support simultaneous visualization of the same 3D

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 202–214, 2018.
https://doi.org/10.1007/978-3-319-91581-4_15

holographic object by multiple participants. Billinghurst et al. in 2000 explored future computing environments research using augmented reality as an interface for collaborative computing [2]. Microsoft HoloLens may be the missing hardware that is needed to make shared augmented reality workspace a reality.

Figure 1 illustrates the view from a Microsoft HoloLens for a mission planning shared augmented reality application where multiple users of Microsoft HoloLens collaborated on a mission planning scenario. Using the shared augmented reality application, users are able to collaborate on a mission plan by manipulating the same 3D holographic object of the aerial map displayed on their respective augmented reality display device while going through the mission objectives. Reducing registration error to ensure proper 3D holographic object placement in such a scenario is crucial to ensure all participants' views are properly synchronized.

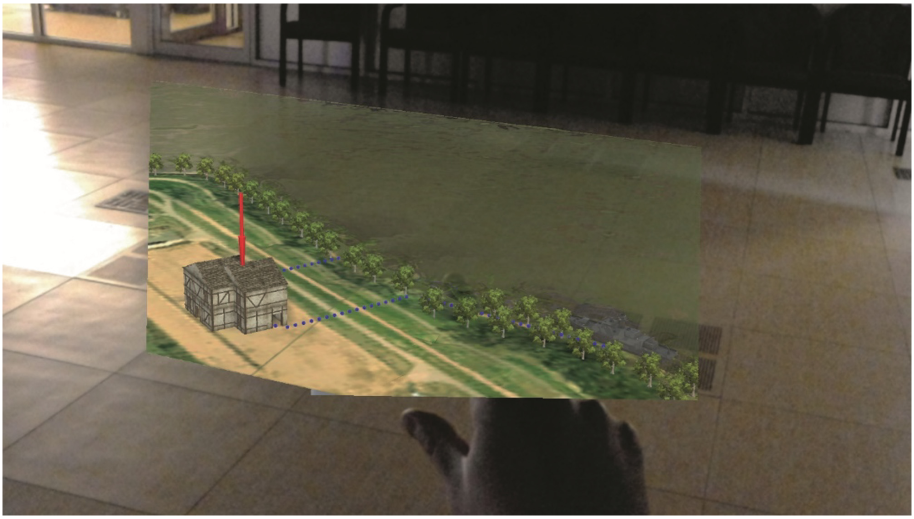


Fig. 1. Using augmented reality for scenario planning. A view from one of the Microsoft HoloLens showing landing and extraction route.

Our experience with existing shared augmented reality demo application using HoloToolkit Sharing Service provided by Microsoft has indicated a noticeable differential in the 3D holographic object placement when the same object is viewed from different Microsoft HoloLens users. 3D holographic object registration is only accurate for the Microsoft HoloLens user doing the placement. For other Microsoft HoloLens users, the same 3D holographic object seemed to be slightly misplaced in the actual environment, resulting in less than ideal shared augmented reality experience. However, since Microsoft HoloLens is mainly a single user augmented reality device, a more accurately placed 3D holographic object for a shared augmented reality application may require additional 3D positioning information that is not available from the array of sensors incorporated into existing Microsoft HoloLens. Our framework used the OptiTrack motion capture system to add positional tracking data of the different Microsoft

HoloLens users to improve the 3D holographic object registration in a shared augmented reality application. The main contribution of this paper is a Unity based sensor data fusion framework that reduces 3D holographic object registration error for a shared augmented reality application and demonstrated the application of the framework using our shared augmented reality mission planning scenario application.

In the remainder of the paper, the related work section discusses the significance of 3D holographic object registration for augmented reality technology. Section 3 describes our sensor data fusion framework and the comparative study design and results showing an improved 3D holographic object registration using our sensor data fusion framework. Then we discuss our shared augmented reality application implementation for a collaborative mission planning scenario in Sect. 4, and we conclude in Sect. 5.

2 Related Work

3D holographic object registration is one of the main challenges of the augmented reality technology. Azuma in his 1997 augmented reality survey paper devoted a whole section to registration challenges [3]. Zhou's et al. survey paper published in 2008 on the trends in augmented reality research also mentioned a significant number of publications in augmented reality are related to registration research [4]. Zhou's et al. paper also has a section on the use of hybrid tracking technique to improve overall tracking and 3D holographic registration accuracy for augmented reality applications.

Hybrid tracking technique combines tracking data from various 3D tracking systems in order to improve the overall tracking quality. Arth used sensors on a modern smart phone to improve 6-degree-of-freedom localizations in wide-area environments [5]. State used landmark tracking and magnetic tracking to provide superior augmented reality registration [6]. You used a hybrid of inertial and vision tracking to improve registration for an augmented reality application [7]. Although Zhou's survey paper mentioned a significant portion of augmented reality research devoted to 3D holographic object registration, there are not a lot of publications targeting registration accuracy research for a shared augmented reality experience.

3 Sensor Data Fusion Framework

In our research to enable collaborative visualization, we addressed the issue of registration inaccuracy of 3D holographic objects in a shared augmented reality environment running on Microsoft HoloLens devices. In registering 3D holographic objects in the real world, Microsoft HoloLens uses their camera sensor inputs to process and generate coordinate systems needed for 3D holographic object placement. For a single user augmented reality experience, the object registration inaccuracy will only affect the view of the single user and there is no requirement to synchronize the views of all users which is required for a shared augmented reality experience. For a shared augmented reality experience with multiple devices, the goal is to minimize the registration error of the 3D holographic object so that all devices see an identical augmented reality world. Our sensor data fusion framework proposes a solution to minimize registration error by fusing external sensor data with the Microsoft HoloLens device's sensor data to improve the registration accuracy for the multiple devices running a shared augmented reality application.

Figure 2 shows the development setup of our sensor data fusion framework. In our implementation, we used Unity to develop a client-server application, with the server being hosted on a dedicated server machine and the client applications deployed and run on the individual Microsoft HoloLens devices. We used three Microsoft HoloLens devices in our shared augmented reality application to demonstrate a synchronized view from three separate devices. For the external sensor data, we used OptiTrack Motion Capture system that uses the Motive software to convert the data captured by the 6 Infra-Red cameras setup we have into 3D positional and orientation data and broadcast the data over the network. The additional 3D positional and orientation data essentially provides the individual Microsoft HoloLens with its location information within a global coordinate system.

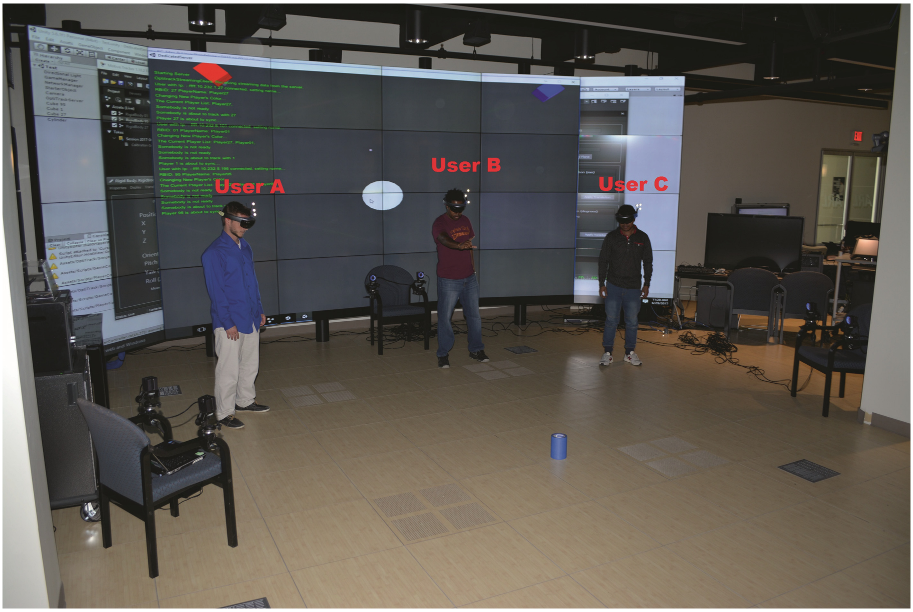


Fig. 2. Our shared augmented reality setup showing relative position of user A, user B, and user C (from left to right) in the real world looking at a 3D holographic object placed on top of the blue object in the real world. (Color figure online)

We imported the global coordinate information into our Unity application using Unity-OptiTrack plugin and combined the data with the 3D positional data from Microsoft HoloLens sensors in our data fusion framework. Since OptiTrack uses marker-based tracking system, we attached markers with unique configuration to each Microsoft HoloLens for Motive to distinguish between the different Microsoft HoloLens devices.

In Unity, we used the 3D positional and orientation data of the Microsoft HoloLens streamed directly from Motive in our server application and fuse it with the sensor data of its corresponding Microsoft HoloLens device running the client application to reduce the registration errors. Having a dedicated server to host the application, the server has the exact information on the physical environment containing the positional and orientation data of each Microsoft HoloLens device and where the holographic object should be rendered with respect to the real world. Thus, each Microsoft HoloLens has a more concise understanding of its position in the global coordinate system and can register the 3D holographic object more accurately with this information. In other words, the server has an exact representation of the world and the location of the holographic objects, and the client application running on Microsoft HoloLens devices merely has to render the scene as defined by the server.

Although we used OptiTrack tracking system to stream the positional and orientation data, the application itself is not dependent on these specific components. Depending on the available sensors and how the data is captured, our data sensor fusion framework can be adapted to incorporate many different types of sensors to further improve tracking accuracy and 3D holographic objects registration. Using the sensor data processed by the data sensor fusion framework, our shared augmented reality application will be able to take advantage of multiple external sensors of different types to improve the object registration inaccuracy.

3.1 Implementation Results and Discussion

Initial use of our shared augmented reality application showed a vast improvement in the accuracy of the holographic object placement for all users compared to the holographic object placement in the shared augmented reality application based on Microsoft HoloToolkit Sharing Service. Computational scientists using the shared augmented reality application can now meaningfully collaborate on a visualization of the 3D holographic object like any objects in the physical world visible from the augmented reality device. When discussing certain characteristics of the 3D holographic object, scientists will be confident that they have a synchronized view.

Figure 3 shows the view from three different Microsoft HoloLens devices. In this shared augmented reality application, we display a fuel injection simulation data [8] where the scientists are interested in the design of the fuel injection subsystem. Since we do not have access to the actual fuel injection hardware at this point, we used a blue physical object in place of a fuel injector. The views are taken from the users A, B, and C standing around the holographic object as shown in Fig. 2. Just as the users are standing around a physical object, the view of the holographic object differ slightly based on the viewing angle from the Microsoft HoloLens. However, all the views are showing the same 3D holographic object being placed on top of the blue physical object in the real world.

Fig. 3. Views from three Microsoft HoloLens showing the placement of 3D holographic object on top of a blue cylinder in the actual environment. From left to right, User A, User B, and User C respectively as viewed from their position as shown in Fig. 2. (Color figure online)

Although our setup has limited the roaming area of a Microsoft HoloLens to the tracking coverage of the OptiTrack, we believe the improved 3D holographic object registration ability contributes more to the usability of the shared augmented reality application. However, we can also expand the 3D tracking coverage if a larger area is needed for the shared augmented reality application by using a different type of sensor with more aerial coverage. Depending on the task requirement, if greater holographic object placement accuracy is needed, we can always use a 3D positioning tracking system with higher precision.

3.2 Comparative Study on Registration Error of Shared Augmented Reality Experiences: HoloToolkit Sharing Service Vs. OptiTrack Sensor Data Fusion

The main purpose of this comparative study is to determine if the shared augmented reality application based on our data fusion framework has less registration error compared to the shared application based on the HoloToolkit Sharing Service.

Comparative Study Design. In our study, we used a completely randomized experiment design with replication and counterbalancing. Our study consisted of two factors and two treatments. We collected data from six male and 2 female participants from our research laboratory, and only one of them had prior experience with augmented reality.

We divided the participants into four groups of two and applied a single treatment to each group. Each treatment contained two factors.

The two factors were the HoloToolkit Sharing Service shared experience, which we assigned as setup number 1, and the OptiTrack Sensor Data Fusion shared experience, which we assigned as setup number 2. For the treatments, the participants either experienced setup 1 followed by setup 2, or setup 2 followed by setup 1. Of the four groups, two of the groups experienced the HoloToolkit Sharing Service first, then the OptiTrack sensor sharing. The other two groups experienced the OptiTrack sensor sharing first, followed by the HoloToolkit Sharing Service.

For the experiment, the goal was to test how well an object registered in a shared user environment. Using the HoloToolkit Sharing Service, the first participant to connect saw the object first, then the second participant would join the shared augmented reality experience and sync to the first via the sharing service. Using the OptiTrack sensor sharing, the first participant would see the object, then the second participant to join would see the holographic object via the OptiTrack sensor's communication of where the HoloLens device was in the physical world.

For both factors, the same holographic object of a fuel spray model was used. Once both HoloLens devices were connected to the shared experience, one participant was asked to place a physical object beneath the holographic object to mark where the object registered from his or her HoloLens' point of view. The other participant was encouraged to give verbal feedback of how close the physical marker was to the holographic object from his or her own device's point of view. The second participant was then given the physical marker and was asked to place it underneath the holographic object from his or her point of view. This gave both participants the ability to see where the other participant's holographic object registered in the real world.

Comparative Study Results. For the evaluation, the participants were encouraged to discuss among themselves on how near or far the holographic object appeared from the other participant's. During the treatment, participants were asked to answer a short survey of questions following each factor, with each set of questions pertaining to that particular factor's experience. The questions were the same for both factors. However, the results varied depending on which treatment the group of participants were assigned, determining which experience they encountered first.

The main question participants were asked was how close the holographic object showed up to the physical marker within the shared experience. The participants were asked to give a rating for how close the holographic object appeared to the physical object in the real world using a Likert scale from 1 to 5, with 1 being 'not very close' and 5 being 'very close'. Each pair of participants was assigned a Group ID letter so that we could keep track of which group experienced which treatment. So for each group, we averaged the rating for each factor as seen by the chart in Fig. 4.

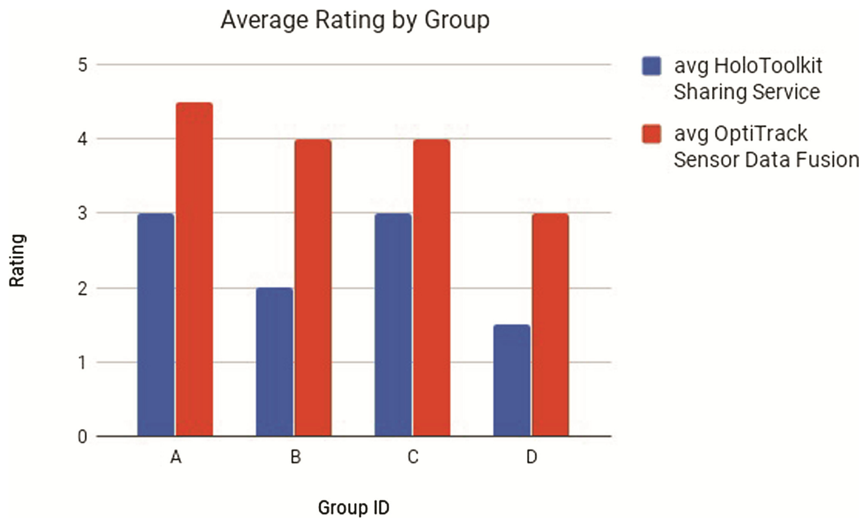


Fig. 4. Results from the comparative study showing average rating by group.

As shown in Fig. 4, the average rating of each group rated the OptiTrack experience higher than the HoloToolkit experience. The most interesting result of the entire study, however, is the average ratings of all groups that took part in similar treatments. There were two separate treatments of the study. The first treatment was having the participants experience the HoloToolkit Sharing Service first, followed by the OptiTrack Sensor Data Fusion. The second experiment was having the users experience the OptiTrack Sensor Data Fusion first, followed by the HoloToolkit Sharing Service. We use setup 1 to represent the Sharing Service experience and setup 2 to represent the OptiTrack experience.

The main take away of the study is the rating given to the second setup the participant's experienced, given the rating they gave to the first factor within their treatment. As shown in Fig. 5, the average rating of both groups rated the OptiTrack experience higher than the HoloToolkit experience in the holographic object placement. For participants that experienced the HoloToolkit Sharing Service first, the rating for that registration averaged a 3 considering the participants had no prior knowledge of sharing registration. When those participants then experienced the OptiTrack Sensor Data Fusion experience following the HoloToolkit experience, the average rating was 4.25 for the OptiTrack. This shows that the OptiTrack experience seemed to do better with registering the shared holographic object. This was a positive reinforcement that the OptiTrack Data Sensor Fusion performs better than the HoloToolkit Sharing Service. On the other hand, when the participants experienced the OptiTrack first, the average rating was only 3.5 for the registration of the holographic object. However, the average rating for the HoloToolkit experience then dropped to 1.75. This shows that having no prior experience, the OptiTrack experience still seemed to look appealing to users with how close the object registered, but also that it definitely did a better job than the HoloToolkit considering the significant drop off in scores for the second experience.

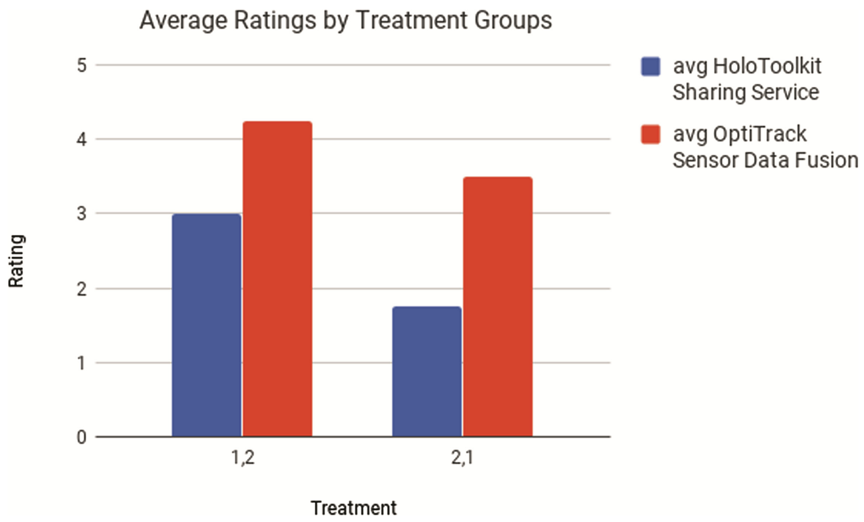


Fig. 5. Results from the comparative study showing average ratings by treatment groups

When analyzing the results of this study, it is more important to observe the average ratings for each treatment as a whole rather than to directly observe the rating from 1 to 5 for the individual factors. As shown in Fig. 5, comparing the average ratings of one factor to the other given what treatment group the participant was assigned shows that using the OptiTrack Sensor Data Fusion shared experience does better to register holographic objects in a shared augmented reality environment. We do not claim to have solved the registration error in shared augmented reality experiences, but rather that we have created a better solution to registering holographic objects using multiple sensors.

4 Shared Augmented Reality Application to Support Mission Planning Scenario

Our shared augmented reality application supports mission planning scenario for multiple users to collaborate in a shared mixed reality environment. The main objective of the application is to enable the users to view and discuss key mission objectives and plans using augmented reality technology such that collaborators can not only view the same 3D holographic scene, but also interact with each other face to face while making eye contact. The mission planning application demonstrates an extraction scenario of a High Value Target (HVT) from an overrun building, with landing and extraction occurring using a sea route (ship). The team of soldiers arrives by ship and assembles at Assembly Area Alpha. The team then follows the predefined path behind the tree line and side of the building to Assault Position Bravo. The team then enters Objective Point through the side door of the compound and takes the shortest path to retrieve the HVT while avoiding patrolling hostiles. Once the HVT has been rescued, the team of soldiers exits the building through the back door, then makes their way to the extraction location. The team will board the ship at the extraction location and the scenario completes. Figure 6 shows the planned path denoted by blue dots showing landing and extraction path.

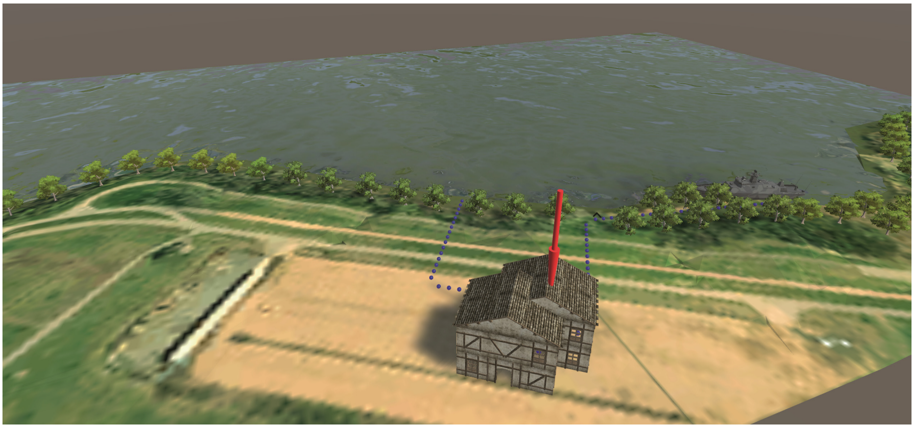


Fig. 6. Planned path in blue showing landing and extraction path. (Color figure online)

The scenario is created in Unity using 3D assets and deployed to a Microsoft HoloLens device. While the building, terrain, and ship assets are all 3D models imported into the scene, the rest of the assets described in the scenario are created using 3D marker within Unity. Currently, the team of soldiers is represented by a blue diamond. Diamonds of different colors can be used to illustrate multiple teams engaged in the mission. The hostiles in the scene are represented by red diamonds that navigate along a fixed path both inside and outside of the compound to simulate aerial patrol. The number of hostiles may be changed and their paths can be manipulated depending on the mission planning scenario. The predetermined path from the landing point, through the building to the HVT, then to the extraction location is represented by a blue dotted line. As the team moves along the path, we illustrate the animation by drawing a solid line along the planned path. The HVT in the scenario is made obvious to the user by a large 3D arrow pointing down toward the location within the building. The arrow is initially grey when the scenario begins, then turns green once the team has successfully rescued it by passing by its location within the building. The extraction point is where the team meets the boat to leave the hostile environment and return to a secure location with the HVT.

Although the application currently runs with this specified scenario, there are many parameters that can be customized in Unity before the application is deployed to the HoloLens device. Depending on the planning scenario, the following parameters can be modified: the number of hostiles, where the hostiles are in the scene, the path of the hostiles, the number of team members, the team's path from landing to extraction, how quickly the team moves along the path, how quickly the ship moves to the extraction point, and the colors of the team, hostiles, path, and target.

4.1 Stand Alone Application

In the stand-alone version of the Mission Planning Scenario application we created, the user is able to run a predefined scenario in a HoloLens device. On launch, a terrain model with a building and a ship placed on it will be shown. To the left of the terrain there is an instruction menu to help the user manipulate the scenario. On the instruction menu is a list of voice commands that the user can say in order to initiate different capabilities within the scene. The user can view the scene from different perspectives by physically walking around in the real-world environment as the scene remains fixed.

However, the user may also manipulate the scene by rotating, dragging, or resizing the scene by using the voice commands "Rotate Scene," "Drag Scene," or "Resize Scene" respectively. To rotate the scene, the user taps and holds, then moves the hand to the left or right and releases the taps to complete the rotation. To drag the scene, the user must also grab and hold, but can then move the hand in any direction. The scene will be placed at the new location where the user releases the tapped press. To resize the scene, the user must tap and hold, then move his/her hand up or down: up to make the scene bigger and down to make the scene smaller. The user's voice command will initiate the specified manipulation technique until the user performs the corresponding action and then releases.

When the scene loads, many aspects of the scene can be toggled on or off depending on what the user wants to focus on. The voice command "Show Team" will toggle on viewing of the team that will be performing the rescue. This is currently a blue diamond that appears next to the boat at Assembly Area Alpha. The "Show Path" voice command toggles on or off the predefined path that the team will take starting from the ship into the building to rescue the HVT, and then to the extraction point. The path is a dotted line that changes to a solid line as the team moves from dot to dot along the path.

The "Show Target" voice command toggles on or off the HVT to be rescued by the team. The target will be a grey arrow pointing toward a location inside the building until the team rescues it, turning the arrow green. The "Show Hostiles" voice command will toggle on or off the hostiles in the scene. The "Run Simulation" voice command plays the animation of the scenario. As the team travels along the path, the hostiles will be moving back and forth, and the boat will make its way to the extraction point. The "Pause Simulation" voice command will pause the scene at any point during the simulation. This allows users to pause the scenario, manipulate the scene by toggling on or off something to view or by rotating, dragging, or resizing, then continuing to run the simulation. The last voice command is "Reset Units" which resets the entire scenario back to the beginning.

4.2 Shared Augmented Reality Application

In the shared version of our application, we have a server application running on the same machine that hosts the OptiTrack's Motive software, and a client application that runs on the HoloLens devices. In order to run the application, the server must first be running and connected to the Motive software data streaming to receive the 3D positional tracking data. Then, the client application running on each independent HoloLens may be started and the client application will connect to the server over the network.

When the client connects, the user will see the predefined scenario in a shared environment. All users will see the scene in the same position and orientation at a tabletop height in the real-world environment, in the center of the OptiTrack cameras' capture area. Underneath the scene is a holographic cylinder that rests on the ground as if the scene were sitting on a real-world object like a table. This allows users to physically point out objects in the scene and discuss with each other naturally as if they were standing around a table in the real world.

The predefined scenario is set with the team in place, the path mapped out, and the target set. There are no voice commands in the shared application as the users will generally be in a close enough vicinity that a voice command would register in multiple users' HoloLens devices. Instead, any user has the ability to begin the scenario by simply performing an air tap gesture while gazing at the cylinder. The air tap registered by one of the clients sends a message back to the server, which then broadcasts the run simulation command out to all connected clients. The team and boat will begin moving and the scenario will play out in unison for all clients. Users are free to move around in order to inspect the scene from different angles and viewpoints, and to discuss with the other users what they are all seeing.

5 Conclusion and Future Work

We investigated using data from different sensors to improve 3D holographic object registration for a shared augmented reality application in our sensor data fusion framework. We used the technique to design a working shared augmented reality application to support collaborative mission scenario planning.

In a sensor rich environment, various sensor data can be used to derive the necessary information about the 3D environment needed to improve a shared augmented reality experience. Similar to the data fusion work with information from multiple data source to generate situational awareness, our existing framework allows us to use data fusion techniques on data from different sensors to enrich our augmented reality experience. We plan to expand our work to include sensor data from networking devices, motion detection sensors, and multi-spectral cameras. Another research area is to build the algorithm to determine when to update or resynchronize data from which external sensors.

Acknowledgments. This work was supported in part by the DOD High Performance Computing Modernization Program at The Army Research Laboratory (ARL), Department of Defense Supercomputing Resource Center (DSRC).

References

1. Furlan, R.: The future of augmented reality: HoloLens - microsoft's AR headset shines despite rough edges. IEEE Spectr. **53**(6), 21 (2016). https://doi.org/10.1109/MSPEC.2016.7473143
2. Billinghurst, M., Poupyrev, I., Kato, H., May, R.: Mixing realities in shared space: an augmented reality interface for collaborative computing. In: 2000 IEEE International Conference on Multimedia and Expo, ICME 2000. Proceedings, Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532), vol. 3, pp. 1641–1644 (2000). https://doi.org/10.1109/icme.2000.871085
3. Azuma, R.T.: A survey of augmented reality. Presence: Teleoper. Virtual Environ. **6**(4), 355–385 (1997). https://doi.org/10.1162/pres.1997.6.4.355
4. Zhou, F., Duh, H.B.L., Billinghurst, M.: Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 193–202, September 2008. https://doi.org/10.1109/ismar.2008.4637362
5. Arth, C., Mulloni, A., Schmalstieg, D.: Exploiting sensors on mobile phones to improve wide-area localization. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012), pp. 2152–2156, November 2012
6. State, A., Hirota, G., Chen, D.T., Garrett, W.F., Livingston, M.A.: Superior augmented reality registration by integrating landmark tracking and magnetic tracking. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, pp. 429–438. ACM, New York (1996). https://doi.org/10.1145/237170.237282
7. You, S., Neumann, U., Azuma, R.: Hybrid inertial and vision tracking for augmented reality registration. In Proceedings IEEE Virtual Reality (Cat. No. 99CB36316), pp. 260–267, March 1999. https://doi.org/10.1109/vr.1999.756960.author
8. Bravo, L., Kim, D., Ham, F., Su, S.: Computational study of atomization and fuel drop size distributions in high-speed primary breakup. J. At. Sprays (2017)

Using Body Movements for Running in Realistic 3D Map

Adhi Yudana Svarajati(✉) and Jiro Tanaka(✉)

Waseda University, Kitakyushu, Japan
adhiyudanas@gmail.com, jiro@aoni.waseda.jp

Abstract. We developed a running support system using body movements as input in a realistic 3D Map. The users use their body movements to control movements in the 3D map. We used depth-camera sensor to track the user's body joints movement as the user moves. The location changes are tracked real-time and the system calculates the speed, which will be used as speed control in our system. This means when the users run in the real life, they will also feel like running in the system naturally. The user will feel the realistic aspect of our system because the speed changes with the user's speed. Realistic 3D map from Zenrin is used in our system. The map consists of several detailed featured such as traffic signs, train station, and another natural feature such as weather, lightning, and shadow. Therefore, we would like to help the users feel more realistic and have some fun experiences.

In our system, we use a head-mounted display to enhance the user's experience. The users will wear it while running and can see the realistic 3D map environment. We aim to provide immersive experience so the users feel as if going to the real place and keep the users motivation high. Integrating realistic 3D map as part of the system will enrich human experience, keep user motivated, and provide natural environment that is similar with the real world. We hope our system can assist the users as one of possible alternatives running in a virtual environment.

Keywords: Realistic 3D map · Running support system · Body movements

1 Introduction

Realistic 3D map and Head Mounted Display developments open new possibilities for creating interactive and immersive interface that enables user to explore and enjoy new experiences in the virtual world. On the other hand, depth-camera is a well-known device that can locate body joints and recognize body movements based on body joints' location changes. There are several depth-cameras in the market, such as Leap Motion which specialized in detecting hands and finger movements We are interested in full body movements tracking using Kinect depth-camera in a realistic 3D map environment.

Realistic 3D map recently becoming more popular due to the increasing need of virtual environment that is similar with the real world. Zenrin, a map company from Japan, created realistic 3d map of cities in Japan. The Zenrin 3D map includes detailed

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 215–229, 2018.
https://doi.org/10.1007/978-3-319-91581-4_16

information such as traffic and road sign, public spaces, and natural landscapes [1]. Several cities already developed and many more will come, giving opportunities for virtual reality (VR) system development. Virtual reality provides many possibilities for developing interactive and immersive applications. Although the use of VR holds risk, such as VR sickness [2], it certainly provides opportunities for new applications and provides a unique experience to the user.

Exercise is one of the most important aspect of human life. We need to exercise in daily routine to keep our body healthy and fit. There are many varieties of exercise, but among them running is popular because people consider it fun and easy to do. Despite the popularity of running, in this modern time, many people do not exercise enough for maintaining their health. There are lots of reasons for keeping people from doing routine exercise, therefore we aim to create a system that people can use for exercising, especially running, anytime and anywhere they want.

In this paper, we propose a support system for running in a realistic 3D map. The user's body movements will be used as input by mapping it using Kinect depth camera [3]. The 3D map used in this system is a replication of the real environment of a city in Japan. The use of realistic 3D map will keep the user motivated and provide familiarity with the real environment.

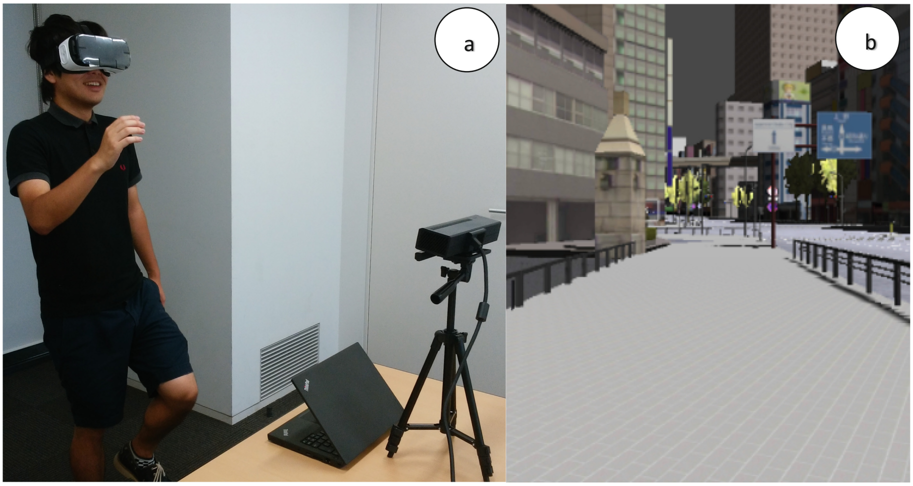


Fig. 1. The users running in front of the camera (a) while looking at the realistic 3D Map through their head-mounted-display (b)

2 Goal and Approach

We are aiming to create a running support system in which the users use his body movement to control movements on the 3D map. When the users run in the real world, they will feel like running in the map naturally. Integrating realistic 3D map will enrich the human experience, keep them motivated, and provide a natural environment.

A system using a realistic 3D map as its environment is still limited, due to the lack of suitable environment. The development of 3D map of cities in Japan by Zenrin provides and opens new ways for presenting immersive experience. Body movement is used as input to the system by mapping the user's body joints into a character in the system [3]. By using body movement as input, the user will feel more connected and realistic in the way they move and control the system. The body joints movement data will be used to calculate velocity and recognize action [4]. The action recognized will determine the movement in the system.

Exploring a large virtual environment usually required lots of space in the real world. Consequently, many systems related to virtual environment exploration could only be realized in small scale or using a controller to move in the virtual world. Using controller to control movement in the virtual world will provide one solution, but consequently the system will not provide natural feeling for the users. Another possible solution is to use walking in place principle to explore a large virtual reality environment, so we do not need to provide lots of space in the real world [5].

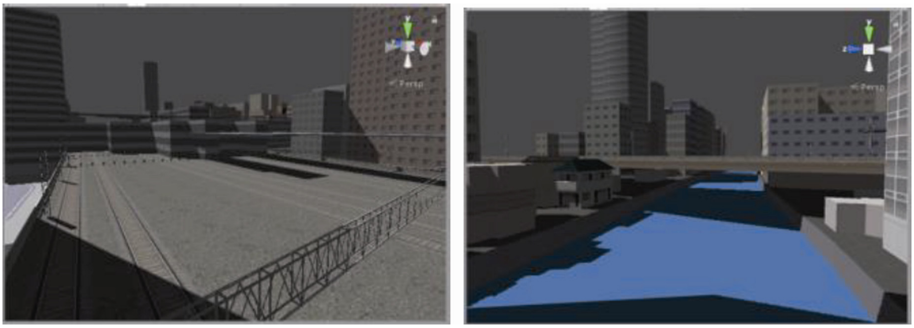


Fig. 2. Realistic environment in Zenrin 3D map

We use 3D map from Zenrin company as our virtual world environment. The 3D map from Zenrin provides detailed and accurate information (Fig. 2) so the user will feel like going to the real place in the real world. The users will also be easier to remember the location and maintain their running performance. They will feel like running in real place, therefore the performance is expected to be similar with the real running.

In this system, we use body movement as input. The user's body movement is tracked and processed into another output in the system. We use body movement to provide natural feeling toward the user, so the user will move at ease and feel like connected into the system through their body. The body movement is represented by body joint's movement in real time.

Among several depth cameras, we choose to use Kinect depth-camera. We choose Kinect as our tool because of its capability to capture whole body data. Kinect can capture the user's body movement through tracking the body joints. Kinect also provides tracking of more than 36 body joints and another feature such as gesture. We would like to use the whole-body movement to produce movement in the system, speed control, posture control, and another feature that require full body movement.

The users will use head-mounted display to look inside the 3D map (Fig. 1a). We decided to use HMD to provide with immersive feeling and natural way. The users can run in longer period and get the feeling of the certain area the user explore in the map.

Our running support system has several merits. Firstly, the users can exercise while looking around the realistic 3D map. This will increase their performance and keep their motivations high. Using body movement instead of another method provides the user connectivity feeling with the system. By using realistic environment, the users can get some insights about the environment and help them remembers the location when later visit the real place in the real world.

3 Running Support System

3.1 Realistic 3D Map

We implemented the system for running support system in a realistic 3D map. We use a 3D map of city in Japan provided by Zenrin. In our system, we use the Otaku city 3D map (Akihabara) because of its popularity worldwide. The 3D map provides a detailed environment and smooth images of the real Akihabara. We downloaded the 3D map file from Unity Asset Store (Fig. 2).

Information in the 3D map consists of several detailed infrastructure, public spaces, and natural landscapes. We can observe traffic and road signs, train station, and rivers just like in the real world. By using this detailed map, we hope can keep the user's motivation high and make the user feel like exploring the location in the real world. The raw file downloaded from Unity asset store still contains several features provided by the company. To develop system of our own, we need to delete the current features and install our own.

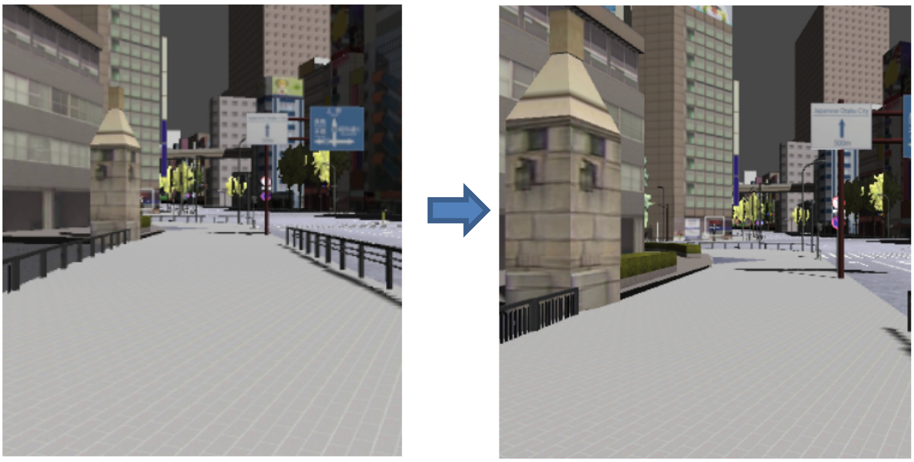


Fig. 3. Point of view changes in the system

3.2 Moving Forward Mechanism

The first feature of the system is the moving forward. The purpose is how to make the user runs as if running in the real world by moving their leg, arms, and create certain posture that are similar. If the users make running like movements, the point of view in the virtual world will change (Fig. 3) as if the users move in the real world. The system will also have speed control for the running movement. Speed in this context means the avatar speed of moving through the map. Figure 4 illustrates the running-like movement performed by the users.

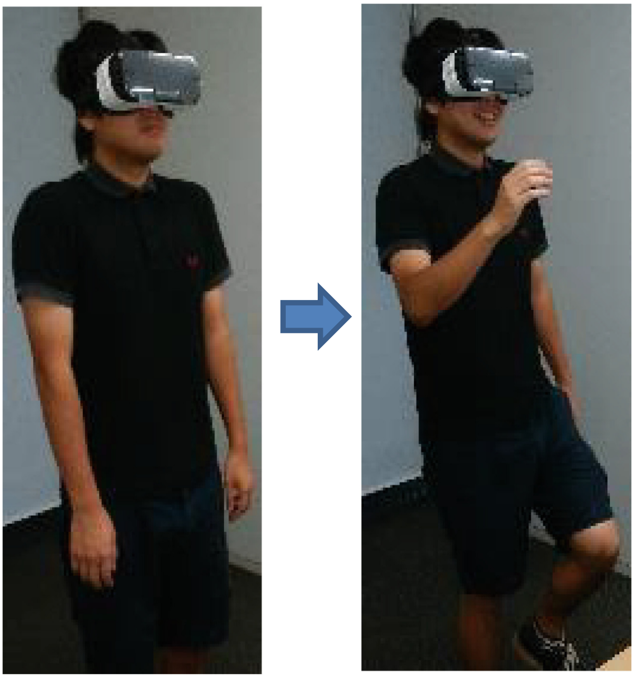


Fig. 4. Running-like movement

First, we use depth-camera to track knee joints. Then, we calculate the approximated speed of knee movements (x, y, z axis). We calculate the velocity by comparing the position changes of those joints for each frame [4] and take the maximum value from right and left knee. The maximum value from both knees movement will be used as the "speed" approximator.

The running speed obtained in our system represents not the whole accurate speed of the user's running activity. The knee movement translated into the steps in our system as approximation of running speed. How many steps we interpreted is flexible and can be changed, but our main point is we want to reflect the change of moving speed in our speed by integrating it with the users running intensity in the real world.

Fig. 5. Captured knee movement

We use knee movement as our input value because of human running posture. In our system, the users do not run like in the real world, instead they will do running in place movement. We observe that knee is the part that will move the most when people running in the real world. Consequently, knee joints are tracked by our system and its changing during the time will be used as speed control input. Figure 5 illustrates the horizontal knee movement we capture to produce the moving forward mechanism and speed control in our system.

Speed controller is added into the system to realize a realistic running support system. Human run in certain speed in certain period and the speed is subject to change due to various reasons. Integrating the speed into the system will provide more realistic feeling to the user, so we can expect linear performances of duration using our system with the benefit of running.

3.3 Horizontal Rotation

Another feature in the system is the horizontal rotation. This feature enables the users to change their point of view in horizontal direction. Since we limit the space of user in our system, we need another way to realize the horizontal movement [8]. In this case, the Horizontal rotation movement is produced by the user's orientation toward the depth-camera.

If the users would like to change the direction in the system, they are required to change direction toward the depth-camera. For example, as shown in Fig. 6(a), if the orientation exceeds x degrees, the point of view in the system will also change toward that direction (Fig. 6(b)). This feature is designed to compensate the vast 3D map environment with the limited space in the real world. By using this feature, the users can change direction and explore another area without changing permanent direction in real life. In this sense, the user will maintain his original position when starting to run again.

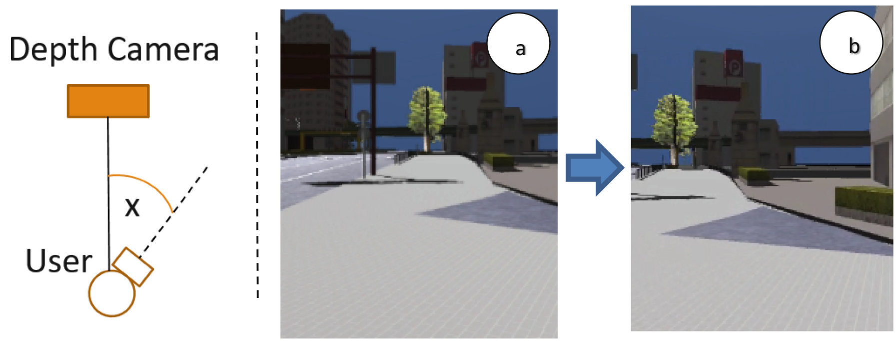


Fig. 6. The user's orientation toward the depth-camera (a) will move the point-of-view in the system (b)

Next, we would like to set certain threshold for control. If the user's orientation change to the right or left side more than the threshold value, we will change the point of view of the system according to user's direction. We set the threshold 45° so the system will not easily detect the user's movement as command for changing direction. When the angle value corresponds with the threshold, the system will trigger the command to the GUI object attached to the avatar. We imbued the GUI object with another command for mouse movement input and limit the moving axis just for x axis. Another axis already covered by the head-mounted display setting for looking around the environment.

3.4 Posture Control

Posture control is a mechanism for comparing the user's posture while running with a standard [9]. The feature will help the user maintain good posture while running, such as asking the users to keep their shoulder and neck straight for maintaining effective running posture (Fig. 7(a)).

In our system the avatar and the users body joints are connected. From that, we can obtain the position of each body joint (x, y, and z) for several body joints such as neck, head, spine, torso, and shoulders. After obtaining the position, we will calculate this following angle: Neck and head angle, torso and neck angle, and angle between shoulders.

We set thresholds within each calculation to keep the angle in certain position so the users will run in good posture. The system will remind the user of their posture by comparing with the standard posture for running. In our system when the user makes bad posture, the system will recognize the posture as bad and reminds the user. For example, in figure the user's running with uneven shoulders position, the system will remind the user by popping a text in the head mounted display as represented in Fig. 7(b).

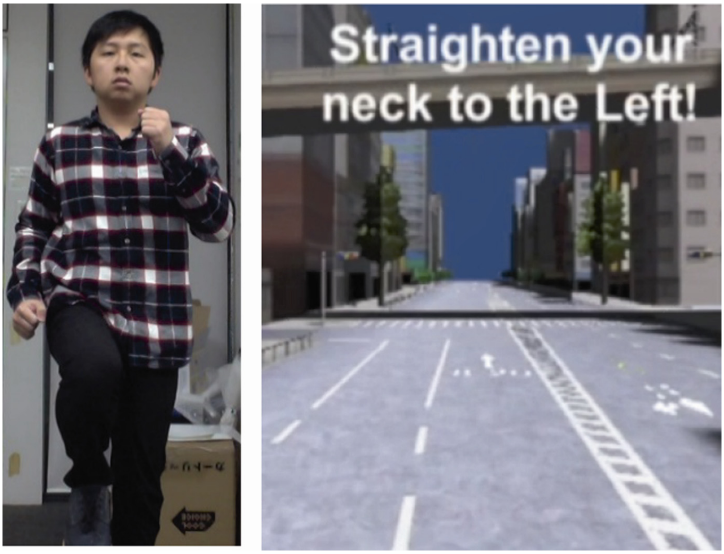


Fig. 7. Good running posture performed by the users (a) and the system reminding the users when they do "bad" running posture (b)

3.5 Collision Control

Collision Control is a feature that will prevents the users to have a collision with another object in the room and makes them stay in the proper distance of Kinect camera. We will use the depth data (z coordinates) of the user's distance from Kinect to their position, and x coordinates data to track the user's movement relative to the sensor in the room.

The system will recognize the distance between the user and another object. In our research, we limited the area of the user movements into a square area of 1.5 m^2 squares in front of the camera. The limitation is meant to prevent the user going out from the Kinect's reading area and prevents the user from making collision or injury by hitting with surrounding objects.

We use the depth camera to extract the distance between user and object (in this case is the table we use for placing the depth-camera). 0.5 m is the maximum distance we allow the users standing in front of the table (Fig. 8(a)). When the users distance with the table decreases below 0.5 m, the system will remind by presenting warning text in the screen. The warning text will disappear when the users move to the safe zone as illustrated in Fig. 8(b).

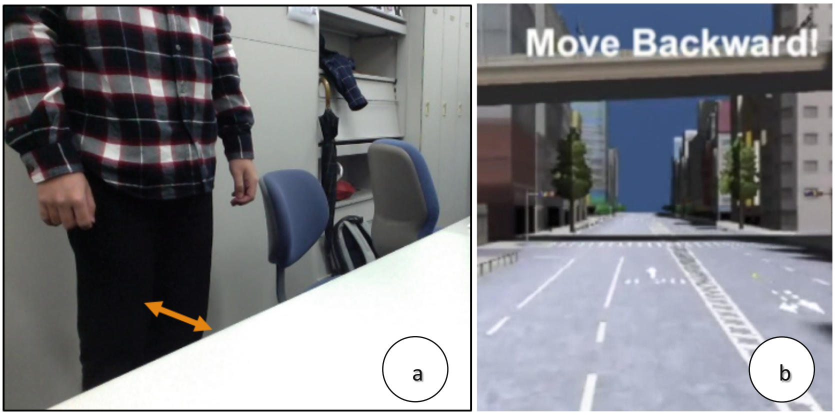


Fig. 8. The User's distance with another object is maintained and the system's warning text when the users move too near with the table

4 Implementation

We use several devices for implementing our system. We choose to use a Kinect depth-camera for capturing full-body skeletal data. We will use Samsung Gear VR as head-mounted display for providing an immersive experience while using our system. We will also use one-unit computer with an Intel i7 processor, 8 GB RAM, CPU speed 2.5 GHz. The system will be integrated with Unity 3D Studio and Microsoft Visual Studio for scripting.

Several SDK will also be used for integrating all the devices. We will use Microsoft Kinect for Windows SDK, Kinect Unity SDK and "Kinect v2 Examples with MS-SDK". After creating the system, we installed it to the HMD by using server-based system that streams the Kinect data to the HMD in real time. Figure 9 represents the system software overview.

We use Kinect SDK for windows to connect the depth-camera with the windows environment. In our system, we use "Kinect v2 Examples with MS-SDK" downloaded from Unity asset store that contains the official Microsoft SDK for ease of integration. Another version for the VR system also used in our system to provide the connectivity between Head Mounted Display with the Kinect. Using server-based system allows the Kinect data to be streamed live into the HMD through internet connection.

4.1 Body Movement as Input

In our research, we use "Kinect v2 Examples with MS-SDK" asset to ease the connection of user's body joints with the avatar. This asset provides us with the avatar designed and integrated with each of body joints recognized by Kinect Microsoft SDK. We import the asset into Unity and set the Kinect in position so the system can recognize the user's body movement. Although the asset provides good integration of body joints and avatar, the avatar only able to move in very limited space with size same in the real life.

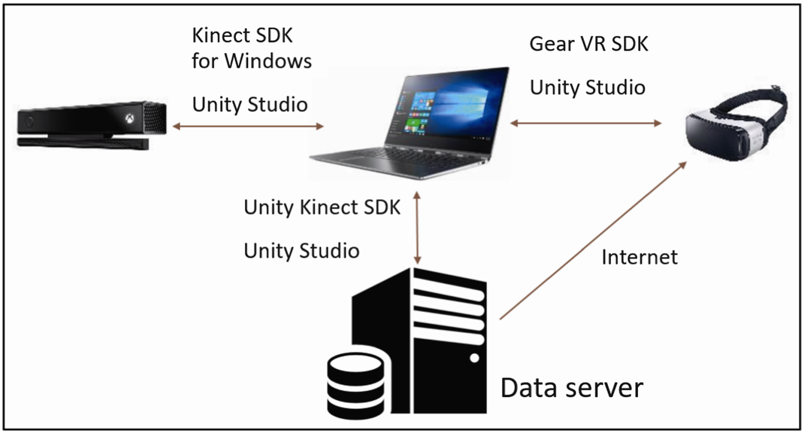


Fig. 9. System's software overview

We connect each joint to a character model in the 3D map using unity and "Kinect v2 Examples with MS-SDK" asset. The user's movement is integrated with the character, as shown in Fig. 10. By integrating the user's movement to the character, we want to make the user feel connected and get "feedback" feeling. "Kinect v2 Examples with MS-SDK" provides the connectivity between an avatar and user, but restricted in movement. Therefore, if we want to move the avatar in a big virtual world, we also need big spaces in the real world [5].

To avoid using too many spaces, we use a "virtual traveling techniques" to cover and travel in such a large distance in the virtual environment [6]. This method allows the user to exploit only small place for their movement. In our research, we implement this method by making the users running in place and the system will recognize it as movement in the virtual environment.

Fig. 10. Body movement integration

One remaining problem is how to make the avatar moves all around the virtual environment. In our research, we will use the basic First-Person Controller movement in the Unity studio to create the basic moving forward mechanism. We will use an GUI object from the Unity that can be moved by using keyboard key. When there is input of that button, the object will move in certain direction, as in our research moving forward. Later we connect the GUI object with our speed control feature, so when the system detects certain speed larger than the threshold, will move the GUI attached to the avatar.

4.2 Body Movement-Based Features

Several features developed in our system such as moving forward mechanism, speed control, horizontal rotation, posture control, and collision control. These features use body movement data obtained from skeletal tracking using Kinect. Figure 11 illustrates the skeletal tracking by using Kinect.

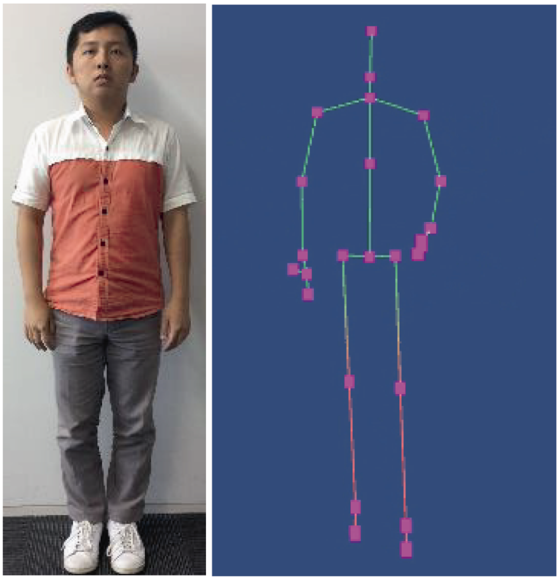


Fig. 11. Body skeletal data captured by Kinect

Speed control uses knee location changes through time, horizontal rotation and posture control uses the users joints locations to calculate several angles, used as threshold value to trigger point of view change in horizontal rotation and warning text in posture control. Collision control uses the depth data (z) as in distance between the users and Kinect camera.

5 Related Work

The use of body movements has been used as method for controlling the VR system in the past. Body movements mentioned in this manner is using body joints or data based on body joints such as joints location changes [3]. The location changes are tracked by certain devices and sensors, some divided per framework and some done in real-time. This method is considered popular and used in various research due to its simplicity and feasible to realize [6].

In the past, realizing this feature in real-time was difficult due of device limitations [3]. However, after the development of Kinect, the users can run the system with proper accuracy, robustness, and affordable. Kinect tracks several body joints and translate it as depth-data, so the users can get the information its positions in the real world [7]. Shotton et al. proposed the principles of Kinect Skeletal Tracker by capturing single input depth image and infer a per-pixel body part distribution [10].

Wilson et al. proposed similar approach by using walking in place method for exploring a large VE [5]. By using two depth cameras and orientation sensors installed above the HMD, they can develop running system. In our research, we use one unit of depth camera without additional sensors.

The use of realistic 3D map in a VR system is implemented for fire evacuation training system [4]. Sookhanaphibarn and Paliyawan use a university based virtual environment in their system, so the users can walk around the area and train themselves regarding evacuation process when a fire emerges.

Comparing with the previous related works, ours hold several advantages. We use minimum devices compared to another works and used simpler configuration and algorithm to create our features. In our work, we use realistic 3D map of Japan city provided by Japan map company. Therefore, the virtual environment presented in our work holds possibilities for another application such as travelling sector. Using realistic environment also enhance the user's experience as the users can use the system while enjoying the environment.

6 Preliminary Evaluation

We conducted a user study by comparing actual running with running using our system. Our hypothesis is by using our system, at least there is 3 major points that the users can take benefit of. Running, enhanced user experience, and similarity with the real environment. While our system is undoubtedly different from the actual running activity, we would like to perceive the differences from the user's perspective, benefits, and further evaluation possibilities.

6.1 Participants and Method

We asked 7 participants (4 Male and 3 Female) to join our user study. All the participants ages are around 23–26 years old, have routine running or jogging schedule during the week, and have never been to the actual location of our system map (Akihabara, Tokyo, Japan).

We asked the participants to do several tasks before. Afterward, we interviewed the participants and asked several questions regarding our system to get some feedback. We would like to let the users experience different environment and compare between those two. The tasks are presented in the following:

1. The participants will run for 10 s using our system
2. The participants will run for 1 min in the street
3. The participants will run for 1 min using our system
4. The participants will run while changing their speed
5. The participants observe the surrounding area while running.

The first task is used to get users' general impressions using our system by running in short period of time. The second and third tasks are for getting the users' impressions of running in longer period and differences between them. The fourth task is for checking the user's impression toward speed control mechanism and the effect it has toward users' performance. The last task is used to get users' experience while using our system.

6.2 Results

After finishing the trial, we will ask the participants to fill a questionnaire. Each question has grade from one to five (1–5; 1 = least agree and 5 = most agree). After filling the questionnaire and sharing some feedbacks, we calculate the average grade for each question. Table 1 presents the questionnaire results.

Table 1. Questionnaire results

Questions		Average grade
Q1	Did you feel that the system intrigues you and enhance your overall experience?	4.71
Q2	Did you find that it is reasonably useful to use the system in longer period?	4
Q3	Did you find that the speed control feature somehow useful to motivate you running in certain speed?	4.14
Q4	Did you find that the horizontal rotation features useful for changing your moving direction?	4
Q5	Did you find that the posture control feature helps you maintain your running posture?	4
Q6	Did you find that the collision control feature helps you to avoid collision with another objects in the room?	4.71

The participants' response toward question 1 is agree and most agree, indicated by the average grade 4.71. The participants feel that by using our system, they can use it not just for running but also looking around the environment. Most of the participants agreed that the realistic environment in our system influences them to repeat and explore the environment more.

Question 2 indicates the participants' attitude toward using our system in longer period. The average grade is 4, which means the participants agree that the system is useful for long usage. However most of the participants also agree that our system can make them losing balance due to repetitive actions and some errors in the system. The participants commented that when the exhaustion due to repetitive actions and the system lagging happens, it slows them down and create some confusions.

The participants' responds toward speed control feature are positive. Most of the participants agree that making the system following their speed makes them to increase their speed. The participants tended to speed up their running motion speed because they wanted to reach certain location faster and explore the area more. Just like running in the real world, when the users' stamina high, they tend to increase their speed, vice versa.

Horizontal rotation feature's usability received average grade 4, which indicates this feature is useful from the participants' point of view. The participants commented that this feature is quite simple and useful for changing the moving direction since they know that the space in the real world is limited. On the other hand, 2 participants stated that this feature makes them a bit uneasy since they need to stop their motion suddenly when they are doing several sprinting motions in short time.

The participants respond toward posture control is positive. They stated that this feature helps to maintain their posture when certain unwanted changes happens (i.e. one shoulder become lower). Some participants even stated that they felt like real running due to the system keeps reminding them to change their posture all the time.

Collision control perceived positively by the participants. They considered this feature important and crucial for preventing injury. All the participants showed some concerns regarding their safety while running using head mounted display before the evaluation began. In the middle of using our system, they could not see their surrounding and this system reminding them to move back to certain direction when they are about to leave the safe zone.

We also received additional comments from the participants regarding the differences of our system with real running. They stated that our system can enhance the running activity until certain extends, for example in a very long period of use, cannot produce the same performance with real running. On the other hand, the realistic environment in our system is the major point as it keeps them wanting to explore and run more. The participants also looking forward to use this system to explore different environments and even interacting with another objects or user in the realistic 3D map environment.

7 Conclusions and Future Work

We designed a running support system in a realistic 3D Map using body movement as input. The user can explore the vast and detailed 3D map while running. We also reflected more natural body movements in our system through speed control, posture control, and another feature in our system. Our system holds possibilities for further improvement and is considered enhancing the users experience.

Based on the evaluation we did with the participants, we can conclude that our system is perceived in positive ways. It indicates that our system effective for motivating the users to run in good manner. Moreover, the users think that exploring the realistic 3D Map environment somewhat improves their motivation for running. Although we use our system for running purpose, it is possible to implement our system in another area such as travelling.

In the future, we are planning to further improve our system by adopting more body movements tracked by Kinect for implementing realistic movement. For example, we can also use the arm movement to interact with objects in the map. Another concern is how to improve the usability of our system when used in longer period due to the risk of losing balance. We are also planning to implement our system with the other platform such as google map or google street view.

References

1. Zenrin 3D Asset. http://www.zenrin.co.jp/product/service/3d/asset/
2. Karina, A., Wada, T., Tsukamoto, K.: Study on VR sickness by virtual reality snowboard. Trans. Virtual Real. Soc. Jpn. **11**(2), 331–338 (2006)
3. Charles, S.: Real-time human movement mapping to a virtual environment. In: Region 10 Symposium (TENSYMP), pp. 150–154. IEEE (2016)
4. Sookhanaphibarn, K., Paliyawan, P.: Virtual reality system for fire evacuation training in a 3D virtual world. In: IEEE 5th Global Conference on Consumer Elevtronics, pp. 1–2 (2016)
5. Wilson, P.T., Nguyen, K., Harris, A., Williams, B.: Walking in place using the Microsoft Kinect to explore a large VE. In: SAP 2013 Proceedings of the ACM Symposium on Applied Perception, pp. 27–33 (2014)
6. Bruder, G., Steninicke, F.: Implementing walking in virtual environments. In: Steinicke, F., Visell, Y., Campos, J., Lécuyer, A. (eds.) Human Walking in Virtual Environments, vol. 10, pp. 221–240. Springer Science+Business Media, New York (2013). https://doi.org/10.1007/978-1-4419-8432-6_10
7. Shotton, J., Fitzgibbon, A., Cook, M., et. al.: Real-time human pose recognition in parts from single depth images. In: CVPR, pp. 1297–1304 (2011)
8. Williams, B., McCaleb, M., et. al.: Torso versus gaze direction to mavigate a VE by walking in place. In: SAP 2013, pp. 67–70. ACM (2013)
9. Tanaka, Y., Hirakawa, M.: Efficient strength training in a virtual world. In: 2016 International Conference on Consumer Electronics-Taiwan, pp. 1–2. IEEE (2016)
10. Taylor, B., Birk, M., Mandryk, R.L., Ivkovic, Z.: Posture training with real-time visual feedback. In: Proceedings CHI 2013 Extended Abstracts on Human Factors in Computing Systems, pp. 3135–3138 (2013)

VRowser: A Virtual Reality Parallel Web Browser

Shuma Toyama[1], Mohammed Al Sada[1,2], and Tatsuo Nakajima[1(✉)]

[1] Department of Computer Science and Engineering, Waseda University, Tokyo, Japan
{toyama,alsada,tatsuo}@dcl.cs.waseda.ac.jp

[2] Department of Computer Science and Engineering, Qatar University, Doha, Qatar

Abstract. In this paper, we propose VRowser, which is a virtual reality (VR) web browser that utilizes various visualization and interaction methods to support webpage content comparisons, allocation, grouping, and retrieval that are essential of parallel web browsing. VRowser's main design objective is to embody a VR parallel web browsing environment that maintains familiar web browsing metaphors and leverages virtual reality interaction and visualization capabilities to support parallel web browsing tasks. Thus, our approach retains the following design factors: (1) Immersive VR. (2) Maintaining document-based web browsing metaphor within the VR environment. (3) Segregated interaction methods for 3D and web-tasks. We present our prototype specifications, followed by an evaluation. Our user study essentially gauged participants' impressions about VRowser, as well as their webpage placement strategies within two different VR environments. The results indicate that users heavily relied on environmental landmarks, such as trees or furniture, to facilitate placement and retrieval of webpages. While the locomotion method developed in our prototype proven to be inefficient for quickly travelling from one location to another. Lastly, we present our conclusion and future development direction of our work.

Keywords: Virtual reality · Parallel web browsing · Information space

1 Introduction and Motivation

Parallel web browsing is becoming prominent part of our daily lives. Nowadays, all users access multiple web sites through web browser tabs or multiple browser windows, whether on tablets, smart phones or personal computers. Yet, tabs and windows are insufficient in support parallel web browsing. Previous work has found that a minimum of 57.4% of users' time is spent switching tabs, where they require continual attention to switch pages [2, 15, 16]. Moreover, the performance of multiple-browser windows in accomplishing parallel web browsing tasks is largely proportional to screen size [25], and overall, they are inferior to interaction techniques that specifically support parallel web browsing [1, 25]. Thus, such shortcomings present multi-faceted challenges in interaction and visualization techniques to successfully support web browsing [24].

As virtual reality (VR) becomes more accessible to people, it is easier to take advantage of the visualization and interaction capabilities enabled by VR. In general,

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 230–244, 2018.
https://doi.org/10.1007/978-3-319-91581-4_17

VR can provide advantages in interaction and visualization that surpass experiences on traditional devices, like desktops and tablets [8]. Yet, such potential remain challenging as the visualization and interaction possibilities are still under heavy researches.

In this work, we propose VRowser, which is a virtual reality web browser that utilizes various visualization and interaction methods to support webpage content comparisons, allocation, grouping, and retrieval that are essential of parallel web browsing [15, 25]. Our approach essentially uses parallel web browsing requirements as determining design factors for VRowser. In this paper, our contributions are the following: (1) VR parallel web browsing system that (A) Maintains familiar web browsing metaphors. (B) Leverages VR interaction and visualization capabilities to support parallel web browsing tasks. (2) Presents evaluation results of our proposed platform that investigates user preferences, behaviors and performance in mentioned parallel web browsing tasks.

This paper is organized as follows. Section 2 presents some related researches to our work. Section 3 describes the design and implementation of VRowser. In Sect. 4, we show the results of the evaluation of VRowser. Finally, Sect. 5 concludes the paper.

2 Related Work

2.1 Web Browsing in Virtual Environments

Researchers introduced numerous metaphors to access and visualize web contents in virtual environments (VE). Some work attempted to alter the document metaphor [4, 5] of webpages, such as representing webpages as rooms in a virtual world [11] or driving a car to navigate through webpages [3, 20].

Moreover, the city metaphor [10, 22] portrays webpages as numerous building, which are textured text and other multimedia information, where users are able to navigate these cities using a variety of locomotion methods. These metaphors could bring numerous advantages for specific web browsing tasks, such as revisitation [22] or navigation [10]. In addition, such collectively exploit associative memory techniques to enable faster retrieval of information at specific locations around the city.

In contrary to work that attempts to utilize various metaphors and interaction methods, additional work within VE attempted to maintain the webpages' document metaphor. The document metaphor is familiar to all users, as it is the metaphor essentially used in all devices to browse the web. Work within this direction mainly textured rectangular 3D objects with webpage contents and integrated various interaction means to deliver the user experience. Earlier work [5, 7] have investigated a method to embed 2D desktop applications within VR, including a web browser, where users could click on various contents using their fingers. Likewise, later works [3, 13] proposed various document-texturing architectures for 3D environments as well as and investigating gesture based interactions with webpages.

2.2 Commercial VR Web Browsing Systems

With the VR revolution that has popularized VR and made it accessible to all users, an expanding number of applications has been made available to experiments with browsing in VR. Standard VR web browsers, such as Oculus browsers[1], intend to provide web browsing experience similar to that of desktops, albeit with modified input capabilities. Additional community developed browsers, such as those found in smartphone VR enclosures, enable users to browse webpages in various locations, whilst maintaining the overall document metaphor. Yet, such browser mimics the interaction and visualization experience on the personal computer, thus, not taking full advantage of what VR can potentially provide.

Lastly, JauntVR [17] alters the visualization of webpages by showing them as rooms with various 3D interactive contents. Despite its innovation, JauntVR requires customizing a webpage's code for compatibility, and we believe the final web browsing experience totally deviates from the standard web browsing experience users are familiar with.

All in all, we argue that such web browsers do not provide significant advantage over desktop web browsers, but rather the primitive capability to view webpages within such VR environments in a similar user experience that is found in desktop computers.

2.3 Web Workspaces

An information workspace [9] is an environment where users can manipulate and store information or documents of interest. VEs embody the concept of workspaces by utilizing varied interaction and visualization metaphors to enhance the users experience, performance or to achieve other objectives within the workspace concept.

Web workspaces has also been previously investigated in various literature. Such work mainly attempted to combine visualization and interaction techniques as a metaphor to interact with webpages and support various tasks. Card et al.'s research provided insights towards interaction in 3D web browsing environments to support revisitation and bookmarking [23]. Their work included; (i) The webbook: a 3D web browsing environment based on the book metaphor; (ii) The web forager: a 3D information workspace which enables users to aggregate various webbooks. Jhavery et al.'s work presented a web workspace that supported a number of web browsing tasks, especially content comparison [18].

Lastly, two-dimensional workspaces have been investigated. For example, Data Mountain [21] supported collecting and grouping webpages a 3D environment, taking advantage of human spatial memory to recall needed webpages.

3 VRowser

VRowser is a VR web browser system that embodies various design factors and interaction concepts to support parallel web browsing. Our approach extends the concept of web workspaces, such as in Data Mountain [21], which takes advantage of spatial memory cues to group and organize webpages in various workspaces. Our work additionally adds the ability to (1) modify page locations in 3D space (2) enable extended

[1] https://www.oculus.com/experiences/gear-vr/1257988667656584/.

viewing space within immersive VR (3) change the size and orientation of webpages within the 3D VR space.

3.1 Design Factors

Immersive VR: The utilization of immersive VR could bring several advantages to the interaction experience. In addition to potential advantages of VEs, such as depth cues [22] and spatial memory [12, 21], VR provides an unlimited view space for users to view and manipulate webpages. Such aspect is significant overcoming limitations of limited screen size [14, 25], or limited workspaces of AR that rely on spatial information registration [1]. Lastly, immersion can be utilized to enhance task performance as in previous work [8, 25]. Moreover, 3D user interface visualization and interaction capabilities can be utilized within VR to enhance the web browsing experience and achieve our design objectives.

Maintaining Webpages' Document Metaphor: As in previous work [1, 21, 25], we have chosen to preserve the document metaphor to insure a familiar user experience with webpages, both in terms of interaction and visualization. Despite the potential advantages metaphors could bring, such as in [6, 7, 14, 19], we believe numerous shortcomings would arise. For instance, other visualization methods, as in the city or gallery metaphors, could have various limitations in terms of restricting amount of viewable contents or longer accessibility cost [6, 10]. Thus, besides users' unfamiliarity and various potential shortcomings, such metaphors require conversion mechanisms that transform webpages from documents to the chosen other forms (As in buildings in the city metaphor, or deconstructed-documents as in the gallery metaphor), which is additionally prone to further interaction or visualization issues. As a result, similar to previous work [3], we have chosen to maintain webpages' interaction and visualization methods.

Segregated Interactions Methods: VRs and webpages require different interaction methods that suite both webpages and VEs. As such, Jankowski [14] classified interaction with webpages within VEs into: (1) Web Tasks: which consist of interactions with webpages, such as scrolling or navigating to a certain URL. (2) 3D Tasks: which cover activities related to 3D-attribute manipulations, such as scaling or locomotion within VEs. As such, we have to provide suitable interaction methods that suit each task type, whilst also fulfilling web activity needs.

3.2 VRowser Concept

VRowser aims to: (1) Provides an immersive web-space experience for webpage management and revisitaiton. (2) Supports content comparisons across webpages. Thus, our approach intends to leverage VR immersion, visualization and interaction techniques to achieve the above objectives.

VR Web Workspace (VRW): Our concept of VRW reflects how users group pages of similar interests, whether in tabs and browser windows while browsing or in different folders for bookmarking. A VRW consists of a virtual space where users can freely manage webpages (Fig. 1). Each VRW contains different geographic layout, such as rivers and mountains, where users can place webpages around. Users can navigate VRW as they do with real world locations. Lastly, users multiple VRWs as required and freely transfer webpages across them.

Fig. 1. VRW House (Upper) and Forest (Lower)

Web page Management is mainly concerned with 3D web tasks, which are mainly *selection, positioning, rotation and scaling* [7]. For *selection,* users ray-cast a hand-beam towards desired webpages to select them. Similarly, selected webpages can be *positioned* by directly pointing the users' hand towards the location desired VRW location. Lastly, selected webpages can be *rotated* and scaled by directly manipulating such attributes using hand gestures; rotating the edge of a page or grabbing both ends of a page and waving-in to scale down and vice versa.

The VR Web Workspace (VRW): Similar to browser windows that comprise tabs, a VRW is a VR space that contains a numerous webpages allocated throughout the space's 3D environment (Fig. 1). Each VRW contains different geographic layout. Users can navigate and create VRW and fill them with situate webpages as needed. Lastly, users can create multiple VRWs as required and freely transfer webpages across them.

Web Page Management: Which consist of *3D web tasks*. For selection, users ray-cast a hand-beam towards desired webpages to select them. Similarly, selected webpages can be positioned by directly pointing the users' hand towards desired VRW location. Rotation and Scaling is done by directly manipulating such attributes using hand gestures; rotating the edge of a page or grabbing both ends of a page and waving-in to scale down and vice versa (Fig. 2).

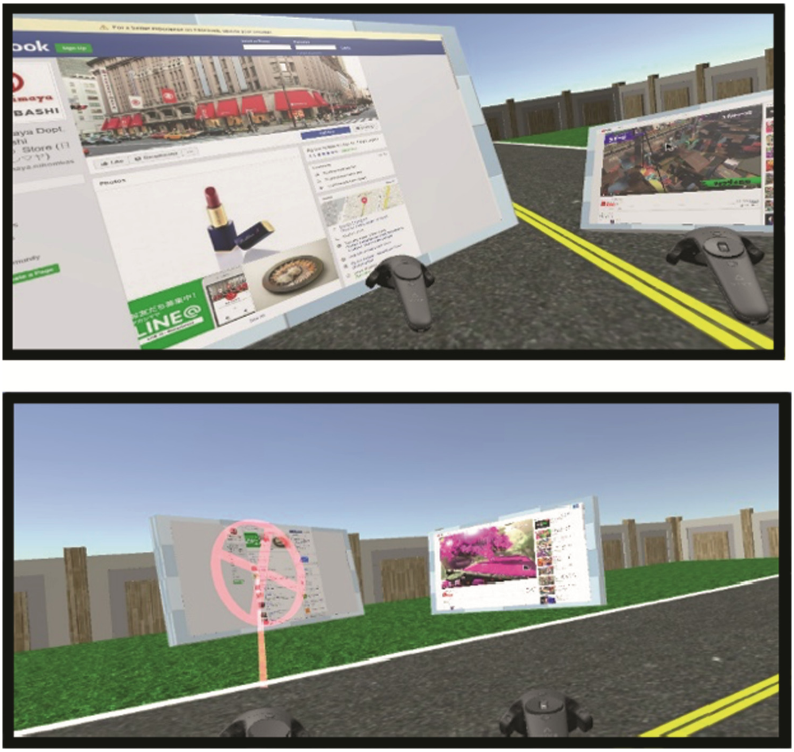


Fig. 2. A user is grabbing and manipulating the 3D location of a webpage (Upper), pointing to interact with web contents (Lower)

Web Tasks: Dedicated inputs and interaction methods are provided to interact with web contents. In Raycasting [7] a user's hands are utilized to control pointers, with a dedicated joystick button to mimic left click (Fig. 2). Likewise, scrolling is done with right clicking and physically dragging in a direction to scroll. Lastly, as a prominent problem in VR, we utilize a keyboard for text entry.

3.3 Implementation

A. *Hardware*

1. **VR Headset:** In this system, we use HTC Vive[2] for VR head mount display.
2. **Interaction space:** Our experience is mainly a seated VR experience, where users are expected sit down while engaging in immersive VR. Yet, the existence of positional tracking, through HTC Vive, it is possible to physically walk around the tracked space and interact with contents.
3. **Interaction Method:** We utilize the HTC Vive's controller as the main method of interaction. Users hold a controller in each hand, where the left controller is essentially used for locomotion tasks within the VRW, while the right controller is utilized to carry web tasks. The possible interactions in each controller are illustrated in Figs. 3 and 4, and include the following:

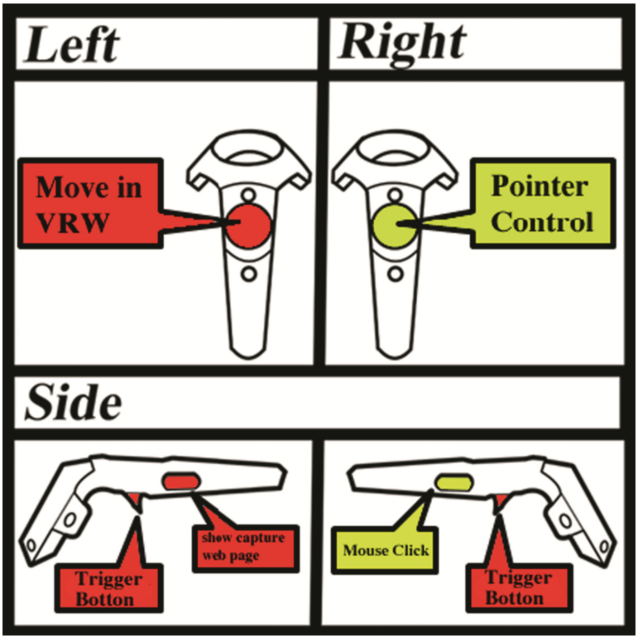


Fig. 3. Operation manual of input device. Green indicates 3D Tasks, Yellow indicates Web Tasks. 3D-Tasks. (The "Trigger Button" has various functionalities effecting 3D Tasks as shown in Fig. 4) (Color figure online)

[2] https://www.vive.com/.

(A) *Scaling*: Upon clicking the Trigger button, users may move the controllers closer or further away from one another to make pages bigger or smaller, respectively (Fig. 4-1).
(B) *Rotation*: Upon clicking the Trigger button next to a webpage (Fig. 4-2), webpages can be rotated by directly rotating the joystick in x, y, z axis.
(C) *Capture and Carry*: this feature allows users to select specific webpages to carry with them, after which they can place it in specific locations within the VRW. To execute this interaction, the user aims the controller towards intended webpage and clicks the "Trigger Button" (Fig. 4-3), after which, (4) Users may view captured webpages by clicking the "Grip Button" (Fig. 4-4). To release a webpage, the clicks the grip-button to view available carried webpages and then points and presses the "Trigger Button" at the desired webpage to be released.

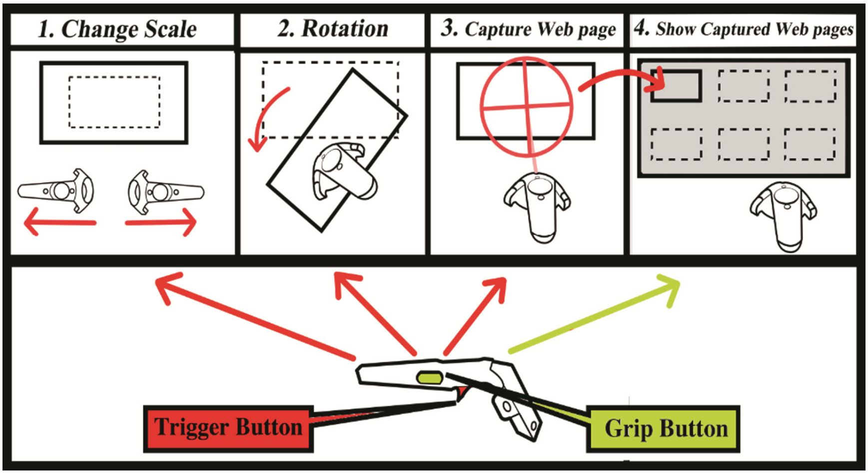


Fig. 4. Description of 3D tasks

(B) *Software*
1. **System:** Our prototype was developed using Unity 3D[3] (version 2017.1.1). We additionally utilize supplementary VR Plugin[4] to use HTC Vive with Unity. Additionally, all input is handled by Unity3D input manager, after which each input is allocated specific 3D or 2D task in accordance with our design factors.
2. **Web Browsing System:** To display web browsers within the 3D environment, we utilize a plug in[5] that enables us to convert webpages to textures, after which we can utilize such textures on 3D objects. Likewise, we developed our control methods to incorporate mentioned plug-in. We access and retrieve webpages using HTTP communication, which allows our system to be expandable for

[3] https://unity3d.com/.
[4] https://www.assetstore.unity3d.com/en/#!/content/32647.
[5] https://zenfulcrum.com/browser.

future iterations that may involve other webpage contents using a similar architecture.

4 Evaluation

4.1 User Study

A. *Design*

Objective: Our goal is to investigate impressions and webpage placement strategies within VRowser

Participants: The experiment was conducted for 10 college students (9 male; 20 to 25 years old) all of whom have had used VR applications before.

Conditions: The user study was carried on two different VRWs, a House and Forest. Each VRW has varied geographical attributes (Fig. 5). While the House includes various rooms and furniture, the forest only includes trees and a simple road. The user study comprised the same set-up described in the implementation section, where we utilized a seated VR experience for this user study (Fig. 6).

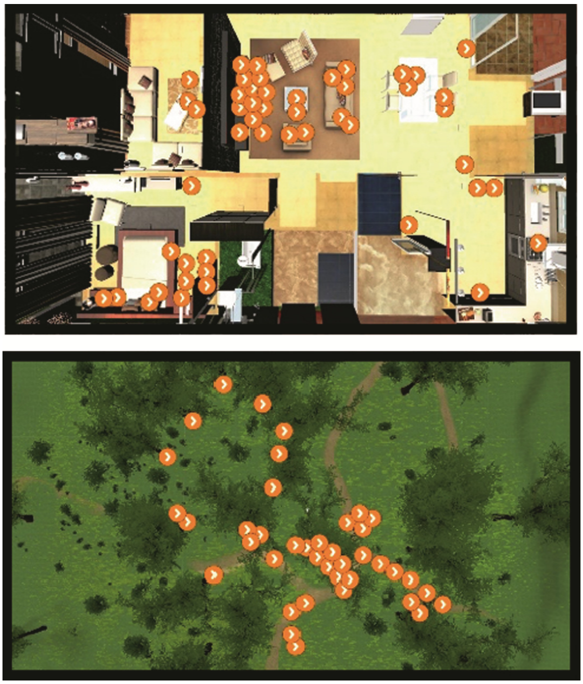


Fig. 5. A heat map that show the concentration of where webpages where allocated each VRW, inside the house (Upper) and the forest (Lower)

Fig. 6. A user wearing the HTC Vive and using VRowser while seated.

Scenario: The study included 10 several genres webpages are arranged in VRW. Thus, after a brief familiarization session, participants were placed within each VRW, starting with the House and then the Forest, and were given 10 webpages to arrange. Next, each participant was given 5 min to check the contents of the webpages, after which they were given 10 min to arrange the pages to their liking within the VRW.

The pages used were the following:

- **Social network service:** Facebook[6], mixi[7]
- **News:** Yahoo! News[8], Dengeki Online[9] (TV game information)
- **Hobbies:** Persona5 Official web site[10], Falcom Official web site[11], comico[12] (A web site to read e-books)
- **Video-sharing:** Youtube[13]
- **Science and Research:** Association for Computing Machinery[14], CHI 2016[15]

6 https://www.facebook.com/.
7 https://mixi.jp/.
8 https://news.yahoo.co.jp/.
9 http://dengekionline.com/.
10 http://persona5.jp/.
11 https://www.falcom.co.jp/.
12 https://www.comico.jp/.
13 https://www.youtube.com/.
14 https://www.acm.org/.
15 https://chi2016.acm.org/.

We diversified the type of webpages to cover a wide range of user interests, and so we can observe how different types of pages may affect the placement at various VRW locations.

Lastly, participants took a questionnaire and a semi-structured interview.

Collected Data: We collected the location of each webpage within the two VRWs. And, after the experiment we interviewed participants about VRowser. The contents of the question asked the degree of satisfaction with the system and which VRW was superior.

B. *Results and analysis*

Web Page Placement

Figure 7 shows the heat map of the placement of webpages in each VRW. In VRW House, many participants placed webpages around featured things in the room. For

Fig. 7. A heat map that show the concentration of where webpages where allocated each VRW, inside the house (Upper) and the forest (Lower)

example, bed is the most prominent object in the bedroom. Also, most webpages were placed in front of the TV in the living room. We thought that this result was generated as participants usually watch webpages on a display. The second most were placed on the bedroom bed. We thought that there were many participants who are interested because the bed stands out from objects in other rooms.

From the interviews, we asked participants about specific page arrangement within VRW. In VRW-House, many participants placed the webpage in association with the characteristics of the VRW. For example, since the bedroom is a private space, most participants placed web site under the category “Hobbies” in the bedroom.

Meanwhile, in the VRW Forest, the placement strategies where quite different from each participant to another. Although participants attempted to utilize a similar strategy as to the House, we believe that the lack of distinguishing map characteristics, such as unique tree arrangements or objects, were a challenge. Thus, participants utilized the scarcity of landmarks by placing many pages on one distinguishing landmark, such as on an easily found tree or where the road forked. Some participant utilized tree condensation as an indicator of amount of information of a placed webpage.

All in all, we conclude that all participants preferred the House VRW over the Forest VRW. Their preference was based on the fact that the house was a place that is easy to navigate in, as it included a variety of objects that acted as landmarks to situate and retrieve placed webpages. In the contrary, the forest lacked sufficient amount of distinct characteristics that helped in navigation or placing webpages.

On the other hand, many participants placed webpages around trees in the VRW Forest. However, as in Fig. 7, the webpage placement was almost randomized and we could not conclude specific placement preferences among users. Based on the above results, we thought that VRW with familiar and noticeable locations and landmarks, like the house, was probably better suited for use as VRW.

Interview Result

Table 1 is a summary of the contents of the interview. First, various participants indicated that the implemented VRW locomotion method was quite slow, elaborating that it could be faster. Thus, we a better locomotion method is required for future iterations, such as warp to arbitrary places quickly.

Table 1. Interview summary

Category	*Question*
System operability	**Were you satisfied with walking in the VRW?** • "*It's good that I don't feel tired even after using for a long time…*" • "*It takes time to move in the VRW (It is slow).*" • "*It requires getting used to..*"
Summary of Participants Comments	**Do you think the 3D Task controls were efficient?** • "*I could directly grab, rotate and move webpages…*" • "*Joystick tracking failed sometimes.*" **Do you think the Web Task Controls were efficient?** • "*....I could use it as if I normally use the mouse on a PC.*"

VRW	**What was your strategy to place the webpage in VRW?**
	• **House**
	  • Associate a webpage and a location. (e.g. Place the Video-sharing web site on or next to the TV set)
	  • Association with everyday real life habits (one participant mentioned "*..I usually check the news on the bed, so I placed the news webpage in the bedroom.*".)
	• **Forest**
	  • "*I placed 2-3 relevant webpages on a tree branch.*"
	  • "*I placed webpages near where the road forked.*"
	  • "*I associated the density of trees with the amount of information on webpages.*"
	Which VRW do you prefer the most?
	• All participants selected House VRW.

Work style	**How would you compare VRowser to standard web browsers?** • "*....Unlike the PC, I could manage webpages by associating them to VRW locations, based on category or other factors.*" • "*Moving a page within the VRW was very time consuming.*"

5 Conclusion

In this paper, we proposed VRowser, which is a virtual reality web browser that utilizes various visualization and interaction methods to support webpage content-comparisons, allocation, grouping, and retrieval that are essential of parallel web browsing. Our approach essentially uses parallel web browsing requirements as determining design factors for VRowser. VRowser shows that virtual reality parallel web browsing system that maintains familiar web browsing metaphors and leverages virtual reality interaction and visualization capabilities to support parallel web browsing tasks.

We believe our evaluation results are promising to pursue further research. Although participants showed strong preference towards VRWs that comprise a larger number of landmarks over those with smaller one, the type of landmarks that a VRW comprise requires further investigation. Moreover, the size of VRW, and how it relates to the number of available landmarks also requires deeper investigations, as such outcome would facilitate creating efficient VR workspaces for future immersive systems.

References

1. AlSada, M., Nakajima, T.: Parallel web browsing in tangible augmented reality environments. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA 2015), pp. 953–958 (2015)
2. Balakrishnan, R.: A study of tabbed browsing among Mozilla Firefox users. In: Proceedings of International Conference on Human Factors in Computing Systems, pp. 673–682 (2010)
3. Barsoum, E., Kuester, F.: Towards adaptive Web scriptable user interfaces for virtual environments. Virtual Real. **12**(1), 55–64 (2008)
4. Benford, S., Snowdon, D., Brown, C., Reynard, G., Ingram, R.: The populated web: browsing, searching and inhabiting the WWW using collaborative virtual environments. In: Howard, S., Hammond, J., Lindgaard, G. (eds.) Human-Computer Interaction INTERACT '97. ITIFIP, pp. 539–546. Springer, Boston (1997). https://doi.org/10.1007/978-0-387-35175-9_81
5. Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I., Brown, C.: VR-VIBE: a virtual environment for cooperative information retrieval. Comput. Graph. Forum **14**(3), 349–360 (1995)
6. Bonnel, N., Cotarmanac'h, A., Morin, A.: Meaning metaphor for visualizing search results. In: Proceedings of the Ninth International Conference on Information Visualization, pp. 467–472 (2005)
7. Bowman, D.A., Hodges, L.F., Bolter, J.: The virtual venue: user-computer interaction in information-rich virtual environments. Presence: Teleoperators Virtual Environ. **7**(5), 478–493 (1998)
8. Bowman, D., Kruijff, E., LaViola Jr., J.J., Poupyrev, I.P.: 3D User Interfaces: Theory and Practice. Addison-Wesley, Boston (2004)
9. Card, S.K., Robertson, G.G., Mackinlay, J.D.: The information visualizer, an information workspace. In: Proceedings of International Conference on Human Factors in Computing Systems, pp. 181–188 (1991)
10. Chiu, P., Girgensohn, A., Lertsithichai, S., Polak, W., Shipman, F.: MediaMetro: browsing multimedia document collections with a 3D city metaphor. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 213–214 (2005)

11. Cibelli, M., Costagliola, G., Polese, G., Tortora, G.: A virtual reality environment for Web browsing. In: Proceedings of the International Conference on Image Analysis and Processing, pp. 1009–1014 (1999)
12. Feiner, S., MacIntyre, B., Hollerer, T., Webster, A.: A touring machine: prototyping 3D mobile augmented reality systems for exploring the urban environment. In: Proceedings of the First International Symposium on Wearable Computers, pp. 74–81 (1997)
13. Gracanin, D., Wright, K.E.: Virtual reality interface for the World Wide Web. In: Proceedings of the 15th Twente Workshop on Language Technology: Interactions in Virtual Worlds, pp. 59–68 (1999)
14. Jankowski, J.: A taskonomy of 3D web use. In: Proceedings of the 16th International Conference on 3D Web Technology (Web3D 2011), pp. 93–100 (2011)
15. Huang, J., White, R.W.: Parallel browsing behavior on the web. In: Proceedings of the 21st ACM Conference on Hypertext and Hypermedia (HT 2010), pp. 13–18 (2010)
16. Huang, J., Lin, T., White, R.W.: No search result left behind: branching behavior with browser tabs. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM 2012), pp. 203–212 (2012)
17. JauntVR (n.d.). https://www.jauntvr.com/. Accessed 22 Jan 2018
18. Jhaveri, N., Räihä, K.-J.: The advantages of a cross-session web workspace. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2005), pp. 1949–1952 (2005)
19. Morris, D., Morris, M.R., Venolia, G.: SearchBar: a search-centric web history for task resumption and information re-finding. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2008, pp. 1207–1216 (2008)
20. Nakaoka, M., Tezuka, T., Tanaka, K.: WebDriving: Web browsing based on a driving metaphor for improved children's e-Learning. In: Bressan, S., Küng, J., Wagner, R. (eds.) DEXA 2006. LNCS, vol. 4080, pp. 499–508. Springer, Heidelberg (2006). https://doi.org/10.1007/11827405_49
21. Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., van Dantzic, M.: Data mountain: using spatial memory for document management. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST 1998), pp. 153–162 (1998)
22. Sparacino, F., Wren, C., Azarbayejani, A., Pentland, A.: Browsing 3-D spaces with 3-D vision: body-driven navigation through the Internet city. In: Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission, pp. 224–231 (2002)
23. Card, S.K., Robertson, G.G., York, W.: The WebBook and the Web Forager: video use scenarios for a World-Wide Web information workspace. In: Conference Companion on Human Factors in Computing Systems (CHI 1996), pp. 416–417 (1996)
24. Thatcher, A.: Web search strategies: the influence of web experience and task type. Inf. Process. Manag. **44**(3), 1308–1329 (2008)
25. Xu, W., Yu, C., Zhao, S., Liu, J., Shi, Y.: Facilitating parallel web browsing through multiple-page view. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 2167–2170 (2013)

Construction of Experimental System SPIDAR-HS for Designing VR Guidelines Based on Physiological Behavior Measurement

Ryuki Tsukikawa[1(✉)], Ryoto Tomita[1], Kanata Nozawa[1], Issei Ohashi[2], Hiroki Horiuchi[2], Kentaro Kotani[2], Daiji Kobayashi[3], Takehiko Yamaguchi[4], Makoto Sato[5], Sakae Yamamoto[1], and Tetsuya Harada[1]

[1] Tokyo University of Science, 6-3-1 Niijuku, Katsushika, Tokyo, Japan
8114067@ed.tus.ac.jp
[2] Kansai University, 3-3-35 Yamate-Cho, Suita, Osaka, Japan
[3] Chitose Institute of Science and Technology, 758-65 Bibi, Chitose, Hokkaido, Japan
[4] Tokyo University of Science, Suwa, 5000-1 Toyohira, Chino, Nagano, Japan
[5] Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, Japan

Abstract. VR technology is still under development and only the evaluation of the performance and functions of VR devices is being studied. There are few studies that are as subjective and difficult to evaluate as the study of immersion in VR. Moreover, methods and indicators for evaluating this characteristic are still unclear. By evaluating this characteristic quantitatively, it seems that technological development for enhancing the immersive feeling in the virtual space will be greatly improved.

Therefore, the final goal of this research is to clarify the relationship between sensory display (force, vision and auditory sense) and Sense of Agency (SoA) in VR environment through physiological behavior measurement. To achieve the final goal, the aim for this paper is to establish a human-scale VR environment and to prepare an environment in which the SoA can be evaluated.

We conducted physiological behavior measurements on two tasks, which are the "ball-catching task" which consists of dropping a ball weighing 220 g from a height of 80 cm and catching it, and the "rod-tracking task" that consists of moving the rod so as not to touch the wall of a sinusoidal path. In the "ball-catching task," it was possible to evaluate the strength of force sense, and the evaluation of the slight force sense was carried out based on the "rod-tracking task."

Keywords: SPIDAR · SoA · VR

1 Introduction

1.1 Background

In recent years, VR has drawn attention not only in the entertainment industry but also in fields such as health care, education, and training. For example, the entertainment industry started out by adopting VR in a game using a head mounted display (hereinafter

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 245–256, 2018.
https://doi.org/10.1007/978-3-319-91581-4_18

referred to as HMD) and now has constructed amusement facilities dedicated to VR [1]. In the medical industry, VR is used for simulation to mitigate risks related to dangerous surgery and rehabilitation [2]. In the field of education, it is becoming possible to carry out education that is easier to understand by using VR that presents temperature and force sense [3]. It is well known that VRs are also actively used in various other fields. Accordingly, people working on the development of VR are also increasing. Many companies and universities are working on a wide variety of VR technologies, such as developing a HMD with a higher image quality [4] and developing devices that can sense force as well as real space in virtual space [5]. We live in an era in which there is a sharp focus on the progress of VR technology.

According to “Virtual Reality Science” supervised by Tachi et al. [6], VR has three elements of “three-dimensional spatiality”, “real-time interaction”, and “self-projection property”. Among these tree elements, “self-projection property” is a feeling of being without conflicts in VR space. It is a subjective concept, therefore, it is hard to evaluate.

This concept has relationship deeply with “Sense of Agency (SoA),” which is an originally conceived research that has been carried out in the field of neuroscience. Studies on the feeling corresponding to a movement of a subject in a VR environment, as if he or she is really manipulating his or her body, have progressed subsequently. Moreover, research that reveals such characteristics is still under way. In such research, the usefulness of applying this idea based on SoA to the field of Neuroscience in the research of VR has been proposed [7]. However, since VR technology is still under development, the performance and functions of VR devices are being studied mainly. Methods and indicators for evaluating this concept are still unclear. By evaluating this concept quantitatively, technological development to enhance the virtual space can greatly improve.

1.2 Purpose and Overview

The final goal of this research is to clarify the relationship between force and vision sense and “Sense of Agency” in a VR environment through measurement of physiological behavior. To achieve this goal, the subgoal of this study is to establish a human-scale VR environment with visual, auditory, and haptics and to prepare a VR environment that can be used to evaluate the “Sense of Agency” through measurements related to performance and physiological behavior.

In this study, we constructed an experimental environment using human-scale haptic device (SPIDAR – HS) and measured physiological behaviors as a preliminary experiment. In this paper, mainly, the construction of the experimental equipment is described.

2 VR System Constructed in this Study

2.1 System Configuration

Figure 1 shows the constructed system. Visual and auditory information is provided using a HTC VIVE, and force information is provided from the end effector of the SPIDAR-HS. The flow of force presentation in this system will be explained by taking as an example where the bottom of a cube held in a hand in Unity is pressed against the ground. First, instructions are issued to wind the motor used in the Unity program. When pushing against the ground, the upper four motors out of the eight motors attached to the SPIDAR - HS are instructed to wind up, so that the cubes held in the hands cannot move further downward. When the motor winds up with the indicated force, the force is transmitted to the thread connecting the end effecter. This is the flow of force sense presentation. In addition, when giving visual information that a cube is pressing against the ground to a person who is already feeling the corresponding force, the person clearly feels that it is pressing against the ground; this is not the case when only the force sense information is provided. In addition, this environmental system must maintain the minimum tension to detect the position angle even when presenting an asthenic sensation. This is to accurately measure the thread winding value by preventing the thread from slacking.

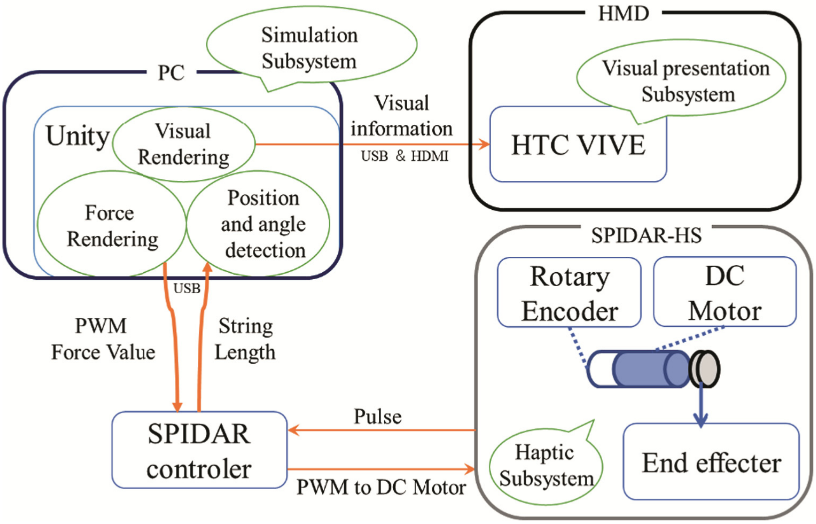


Fig. 1. Configuration of experimental system using SPIDAR-HS

2.2 SPIDAR

SPIDAR-HS. In this study, we employed the SPIDAR system [8–12] as the force presentation device. SPIDAR is a haptic device in which a motor, a threaded pulley, and an encoder for reading a yarn winding are set as one module. In addition, the force sense presentation and position and angle detection are performed by a plurality of motor modules. This SPIDAR system was further expanded to a human scale; this scaled-up system is called the SPIDAR-HS (Human Scale).

2.3 Outline of the Experimental System Using SPIDAR-HS

The construction of SPIDAR-HS is shown in Fig. 2. The size of one side of the aluminum cube is 2.5 m; thus, it is possible to secure a wider working space than the conventional SPIDAR workspace. To use the HTC VIVE for HMD and track the HMD position information on room scale, it is also possible to perform tasks while moving within the 2.5 m cube that we have constructed. Moreover, since the shape of the force sense presentation part (hereinafter referred to as end effector) and the motor attachment position can be changed, it is an experimental environment that can be applied to various physiological behavior measurement experiments by replacing these components according to the application. The specifications of the motor and encoder arranged are shown in Tables 1 and 2. The yarn used is a thread used for fishing, the thickness of which is 0.37 mm. A snap which is a fishing tool is attached to the tip of the yarn. This is installed to connect with the ring created by a fishing tool attached to the end effector body and a quick knot. HTC VIVE is used as HMD. The specifications are shown in Table 3.

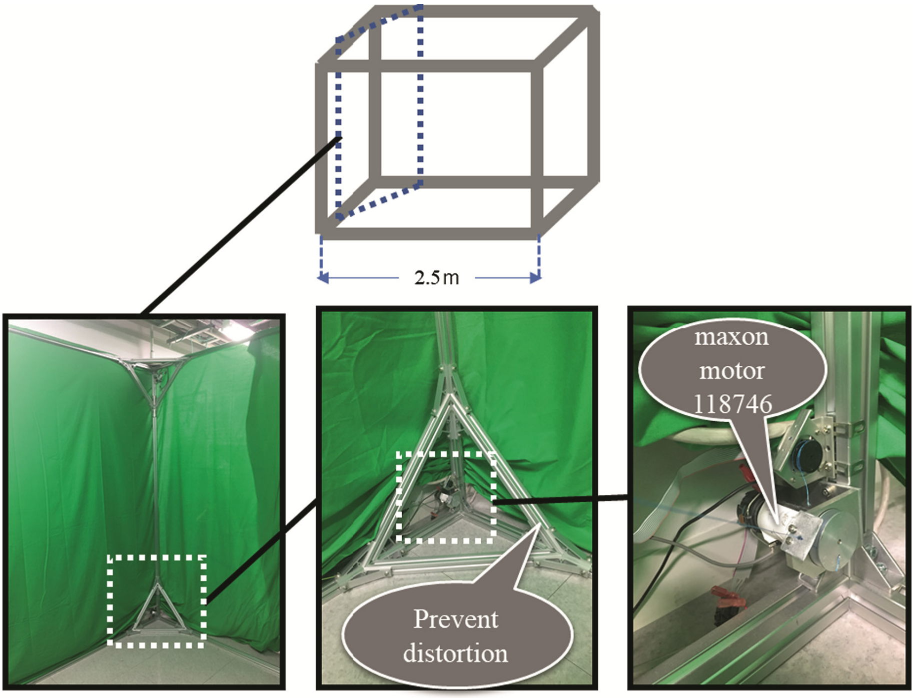


Fig. 2. Construction of SPIDAR-HS

Table 1. Motor specification

Maxon motor 118746 RE25 specification	
Nominal voltage [V]	24
Starting current [A]	3.1
Maximum continuous torque [mNm]	28
Pulley radius [m]	0.01

Table 2. Encoder specification

HEDL 5540 specification	
Count/rotation (resolution)	500
Maximum frequency [kHz]	100
Maximum allowable speed [rpm]	12000

Table 3. HMD specification

HTC VIVE specification	
Screen	dual AMOLED 3.6-in diagonal
Resolution	1080 × 1200 pixels per eye
Refresh rate	90 Hz
Field of view	About 110°
Weight	470 g

2.4 Measurement of Position Error of Working Space

Purpose. The purpose of this experiment is to measure the position error in a work space and evaluate the construction environment. The work space was defined as a cube of 1 m side because tasks needed in this fiscal year will be carried out in this space.

Method. We decided the coordinates of the center as (0, 0, 0) and moved slowly from there to one vertex (−0.5, 0.5, 0.5) of the cube defined as the working space. Coordinates of the Unity side were measured ten times after moving it, and the mean error and standard deviation were recorded. The end effector used at this time was the one shown in Fig. 3.

Fig. 3. End effector in this measurement (also used in experiment 2)

Result and Discussion. Table 4 shows the measured coordinates *x*, *y*, *z*, the error average, and standard deviation. From Table 4, it can be said that the error value is small if the working space is a cube with a 1 m side. However, depending on the tasks, even an error of only 0.5 mm may have a big influence. In such a case, we must reduce this error. What can be considered as the cause of the error is how the yarn is wound. The position of the Unity is calculated assuming that the pulley is not threaded. In fact, the yarn is wrapped around the pulley and it becomes thicker by that much, and the amount that can be wound up per revolution differs from what is estimated; this difference produces errors. To reduce this error as much as possible, it is necessary to reduce the winding weight by using a thinner thread in the experiment.

Table 4. Relative error, standard deviation, average value of *x*, *y*, *z*

	x	*y*	*z*
Relative error [%]	0.38	−0.53	0.96
Standard deviation [m]	0.011	0.0038	0.017
Average value [m]	0.498	0.496	0.505

3 Experiment 1: Ball-Catching Task

3.1 Overview

Purpose. The purpose of this experiment is to determine whether there is a difference in the nature of the muscle activity of the forearm between the Real task and the VR task. This research is a collaboration between the Kotani Laboratory, Ergonomics Laboratory, Faculty of Systems Science and Engineering, Kansai University.

Methods. The experimental conditions are shown in Table 5. Under these conditions, experiments were carried out according to the following procedure.

Table 5. Experimental conditions

Experimental conditions	
Ball initial position	Just over the palm 80 cm
Ball mass	220 g
Ball diameter	80 mm
Myoelectric potential measurement location	Palmaris longus muscle Extensor carpi radialis longus muscle

The procedure of the experiment is as follows.

1. Measurement of resting electromyograms of the elongated extensor carpi radialis longus muscle and palmaris longus muscle for 15 s.
2. Subjects wear the HMD and catch the ball dropped from a height of 80 cm in the VR environment.
3. Measurement of the electromyogram when the subject catches the ball.

The task of catching the ball was performed three times. Next, HMD was removed and the ball-catching task in the Real environment was performed three times using the same procedure that was used in the VR environment.

3.2 Implementation

End Effector. Figure 4 shows the end effector used in the experiment. The frame part is made of resin made with a 3D printer. This end effector is a hemispherical Styrofoam which is of the same size as the actual sphere attached to the lower part of the frame.

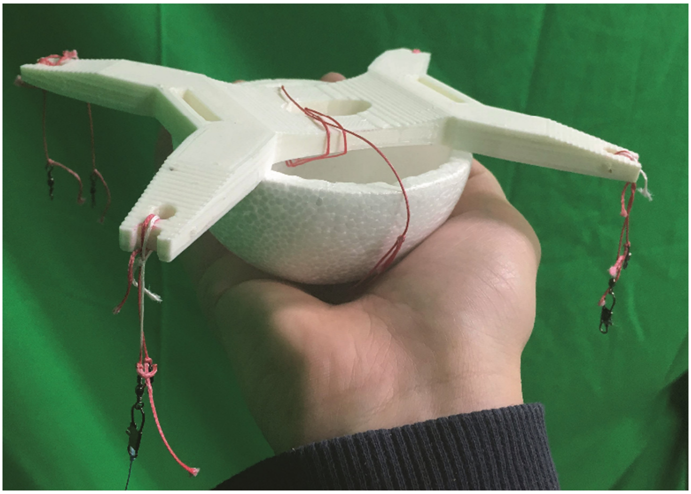


Fig. 4. End effector used in the ball-catching task

Force Sense Presentation. In this experiment, it was assumed that the falling ball was caught in 0.1 s, and the force sense was presented vertically downward for 0.1 s.

VR Environment. The hand and ball in the VR task are shown in Fig. 5. The hand and ball in the Real task are shown in Fig. 6. The colors of the balls and hands in the VR task are nearly the same as those in the Real task. However, the hand in the VR task cannot move like the hand in the Real task. Also, subjects cannot see their own bodies.

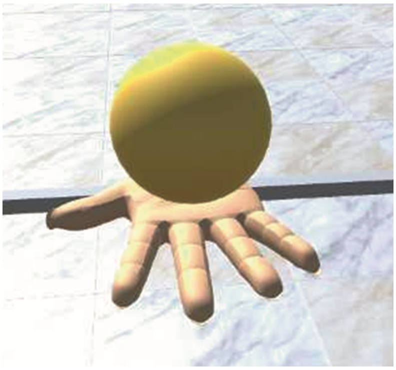


Fig. 5. The hand and ball in the VR task (Color figure online)

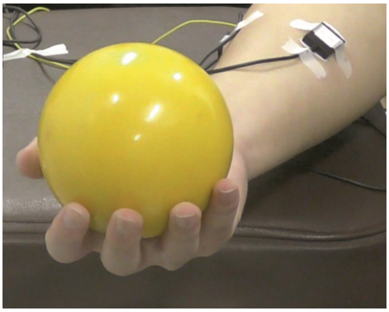


Fig. 6. The hand and ball in the Real task (Color figure online)

3.3 Results and Discussions

The results analyzed by Kansai University, especially the level of muscle activity during catching, indicate that the muscle activity level of the VR task tended to be smaller than the muscle activity level of the Real task. It is inferred that the force presented during the VR task was smaller than the force presented during the Real task.

The impact force applied to the hand when a 220 g ball falls from a height of 80 cm was calculated. The impulse when a 220 g ball falls from 80 cm and contacts the hand is as follows.

$$mv = 0.22 \times \sqrt{2 \times 9.81 \times 0.8}$$

If the falling ball was received in 0.1 s, the required force magnitude F[N] is as follows. An illustration of how to apply force while catching the ball is shown in the Fig. 7.

$$F = \frac{mv}{t} = \frac{0.22 \times \sqrt{2 \times 9.81 \times 0.8}}{0.1} \times 2 \approx 8.18[\mathrm{N}]$$

The maximum presentation power of SPIDAR - HS was measured using a force gauge (RZ-2) and it was 6.23 N. If SPIDAR - HS cannot output a force of at least 8.18 N, it is impossible to measure muscle activity that is exactly like the Real task. It is thought that this is the reason why the level of muscle activity of the VR task becomes clearly smaller than that of the Real task.

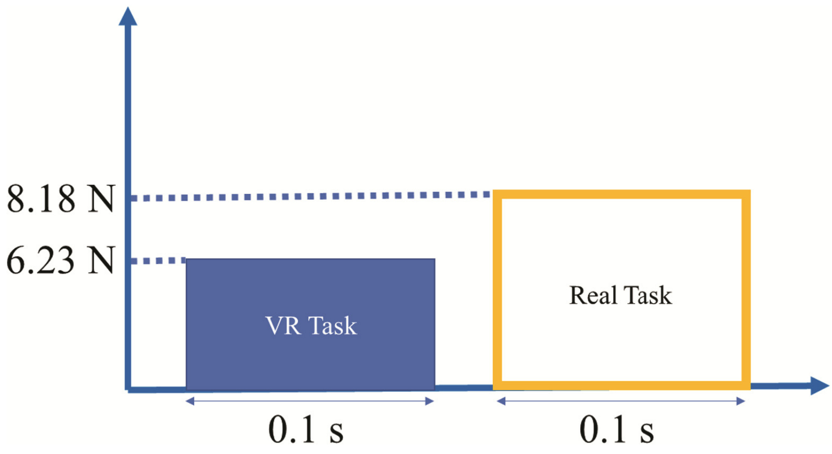


Fig. 7. Force sense presentation while catching the ball

4 Experiment 2: Rod-Tracking Task

4.1 Overview

Purpose. The aim is to clarify the influence of the prototyped task on the sense of motor participation by using EMG in a VR environment. This research is a collaboration with the Kobayashi Laboratory, Chitose University of Science and Technology, Faculty of Science and Technology, and Department of Information Systems Engineering.

Methods. The experimental conditions are shown in Table 6. The minimum required force vertical upward was measured by a force gauge (RZ-2).

Table 6. Rod-tracking task experiment conditions

Experimental conditions	
Rod length	400 [mm]
Rod diameter	10 [mm]
Path size	200 × 300 × 20 [mm]
Path thickness	20 [mm]
Minimum required force vertical upward	0.42 [N] (Maintain constant 6.32 [cm/s])
Myoelectric potential measurement location	Abductor digiti minimi muscle Brachioradialis muscle

The procedure of the experiment is as follows.

1. In order to measure the myoelectric potential at rest, with the right arm pulled out, stop for 10 s.
2. Hold the HMD and grasp the position of the yellow tape on the end effector.

3. With the arms stretched, bring the rod into contact with the start point (far side) of the route; then count 5 s and start.
4. When reaching the return point, count again for 5 s and start again from there.
5. Return to the starting point and count 5 s.

This procedure is carried out until the test subject no longer touches. During the task, sit, so that the body is parallel to the desk and try not to move the body and head as much as possible.

4.2 Implementation

End Effector. The end effecter used is the one shown in Fig. 3. A part made using a 3D printer is attached to the end of a bar of cypress. The yarn from the motor is attached to the part. A vinyl tape is wrapped around the bar holding part.

Force Sense Presentation. When a bar is sunk in an object, the force sense is presented to pull in the direction opposite to the object. The magnitude of the force sense depends on the degree of depression and the speed at the time of sinking.

VR Environment. Figure 8 shows a route in the VR task and a route in the Real task. The left is the route in the VR task and the right is the route in the Real task. The route in the VR task is created referring to the route in the Real task. Since both the route in the VR task and the route in the Real task are the blind spots of the route, it is difficult to see. In the VR task, the subject's own arm and body cannot be seen.

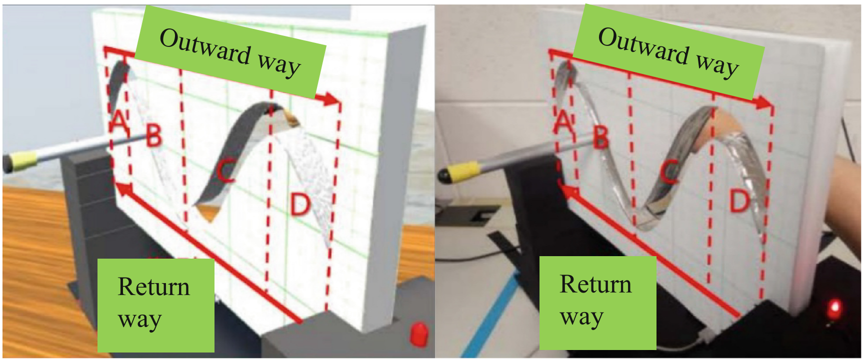


Fig. 8. A route in the VR task (left) and a route in the Real task (right)

4.3 Result and Discussion

Subjects answered that they were unable to recognize which side of the top or bottom the rod is facing. The SPIDAR - HS system applies a force in a certain direction whereby the end effector does not sink into the wall further when the rod contacts the path. The feeling when receiving this force is a feeling that the rod stops rather than feeling of the receiving force. Also, in case of rubbing on the wall of the route or slightly touching it, the force sense decreases. When feeling this small force sense in the absence of visual information, the upper/lower discrimination becomes difficult because the reaction force at object contact cannot be felt so much.

5 Conclusion and Future Work

The purpose of this research was to construct an experimental environment to devise a method for quantitative evaluation of SoA. For this purpose, this year we conducted physiological behavior measurements in cooperation with the ergonomics laboratory of the Faculty of Systems Science and Engineering at Kansai University and the Kobayashi Laboratory of the Faculty of Science and Engineering at Chitose University of Science and Technology and evaluated the construction environment from the results.

The problems of the construction environment that manifested in the ball-catching task were the small forces that could be presented. In the future, depending on the task, it is necessary to solve the small value of the presentation power by changing the motor to a more powerful one, the mounting position of the motor, and the force sense presentation method.

Another problem that was observed is that it is difficult to understand the upper and lower contact in the current SPIDAR system. If physiological behavior measurement is to be performed in tasks requiring a weak force sense in the future, it is necessary to make some countermeasures. For example, it is a mechanism that generates weak vibration on the side when it comes in contact with the end effector. When creating an end effector including such a mechanism, it is necessary to make it smaller and lighter.

In the two tasks performed this time, the subject was not able to see his own hands and body. In the ball-catching task, hands in the virtual space do not move at all; so, it seems that the immersive feeling is lost. Even in the case of the rod-tracking Task, there were also subjects who answered that they could not notice the tilt of the rod or hand that would normally be noticed because their own arms and bodies could not be seen. In the future, we will make our body visible in the virtual space using Chroma Key processing. As a result, when the hand of the subject moves, the hand displayed in the virtual space will move to the same extent and the immersive feeling should be much higher than that observed in the present experiment.

In the experimental environment system constructed using SPIDAR - HS, various end effectors can be attached, and the method of presenting forces can be changed freely. This year, by using different end effectors and force sense presentation methods for each of the two tasks, we were able to obtain results of different physiological behavior measurements. Also, as compared with the conventional SPIDAR series, the work space became very wide by adopting the shape of a cube with a 2.5 m side. This makes it possible to deal with tasks that could not be handled by a conventional SPIDAR. From the above, by making full use of the wide working space, the end effector which can be made into various shapes, and the degree of freedom of the motor mounting position this fiscal year, we can conclude that we were able to construct an experimental environment that can be used to conduct a wide variety of physiological behavior measurements under various conditions in virtual space. From next fiscal year, we plan to increase the performance of the whole experiment system using SPIDAR - HS and to design the VR guidelines for the visual-haptic presentation environment which is the final goal of this research.

Acknowledgement. This work was supported by JSPS KAKENHI Grant Number JP17H01782.

References

1. VR PARK TOKYO. http://www.adores.jp/vrpark/. Accessed 16 Nov 2017
2. FUJITSU. http://www.fujitsu.com/jp/solutions/business-technology/vr-solution/case-study/medical/. Accessed 16 Nov 2017
3. Yamagishi, S., Murayama, J., Hirata, Y., Sato, M., Harada, T.: An experiment-based learning support system for thermodynamics which presents haptic and thermal senses. In: Joint Virtual Reality Conference of ICAT – EGVE – EuroVR (2012)
4. STARVR. https://www.starvr.com/#about. Accessed 06 Nov 2017
5. Nagai, K., Tanoue, S., Akahane, K., Sato, M.: Wearable 6-DoF wrist haptic device "SPIDAR-W'. In: SIGGRAPH Asia 2015 Haptic Media and Contents Design, Article no. 19 (2015)
6. Tachi, S., Sato, M., Hirose, M.: Virtual Reality Sience, 6th edn. Japan Virtual Reality Society, Tokyo (2016). (in Japanese)
7. Kilteni, K., Groten, R., Slater, M.: The sense of embodiment in virtual reality. Presence: Teleoperators Virtual Environ. **21**, 373–387 (2012)
8. Sato, M., Hirata, Y., Kawarada, H.: Space Interface Device for Artificial Reality –SPIDAR-. Shingaku (D), J 74–D 2(7), 887–894, 403–412 (1991). (in Japanese)
9. Hasegawa, S., Inoue, M., Kim, S., Sato, M.: Tension calculation method for tension type force sense presentation device. J. Robot. Soc. Jpn. **22**(5) (2004). (in Japanese)
10. Tashima, H., Furuya, S., Ma, S., Akahane, K., Sato, M.: About Isotropic Force of a 6 DoF Haptic Interface SPIDAR-I. C4 - 2, 1 March 2014
11. Toshishima, M., Akahane, K., Sato, M.: A proposal of haptic interface SPIDAR-S attached to the Smartphone and controlled by audio signal. In: Interaction 2015 Paper collection, pp. 156–159, 26 February 2015. (in Japanese)
12. Isshiki, M., Lin, L., Akahane, K., Sato, S.: SPIDAR-mouse: a development of user generated haptic device. In: Information Processing Society Symposium Proceedings, vol. 4, pp. 55–58, 22 February 2010. (in Japanese)

Augmented, Mixed, and Virtual Reality Enabling of Robot Deixis

Tom Williams(✉), Nhan Tran, Josh Rands, and Neil T. Dantam

Colorado School of Mines, Golden, CO, USA
twilliams@mines.edu

Abstract. When humans interact with each other, they often make use of *deictic gestures* such as pointing to help pick out targets of interest to their conversation. In the field of Human-Robot Interaction, research has repeatedly demonstrated the utility of enabling robots to use such gestures as well. Recent work in augmented, mixed, and virtual reality stands to enable enormous advances in robot deixis, both by allowing robots to gesture in ways that were not previously feasible, and by enabling gesture on robotic platforms and environmental contexts in which gesture was not previously feasible. In this paper, we summarize our own recent work on using augmented, mixed, and virtual-reality techniques to advance the state-of-the-art of robot-generated deixis.

Keywords: Human-robot interaction · Deixis
Nonverbal interaction · Teleoperation

1 Introduction

When humans interact with each other, they often make use of *deictic gestures* [1] such as pointing to help pick out targets of interest to their conversation [2]. In the field of Human-Robot Interaction, many researchers have explored how we might enable *robots* to generate the arm motions necessary to effect these same types of deictic gestures [3–8]. However, a number of challenges remain to be solved if effective robot-generated deictic gestures are to be possible *regardless of morphology and context.* Consider, for example, the following scenario:

A mission commander in an alpine search and rescue scenario instructs an unmanned aerial vehicle (UAV) "Search for survivors behind that fallen tree." The UAV can see three fallen trees and wishes to know which its user means.

This scenario presents at least two challenges. First, there is a problem of morphology. The UAV's lack of arms means that generating deictic gestures may not be physically possible. Second, there is a problem of context. Even if the UAV had an arm with which to gesture, doing so might not be effective; picking out far-off fallen trees within a forest may be extremely difficult using traditional gestures.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 257–275, 2018.
https://doi.org/10.1007/978-3-319-91581-4_19

Recent advances in augmented and mixed reality technologies present the opportunity to address these challenges. Specifically, such technologies enable new forms of deictic gesture for robots with previously problematic morphologies and in previously problematic contexts. For example, in the previous example, if the mission commander were wearing an augmented reality head-mounted display, the UAV may have been able to pick out the fallen trees it wished to disambiguate between by circling them in the mission commander's display while saying "Do you mean *this tree*, *this tree*, or *this tree*?".

While there has been little previous work on using augmented, mixed, or virtual reality techniques for human-robot interaction, this is beginning to change. In March 2018, the first international workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interaction (VAM-HRI) was held at the 2018 international conference on Human-Robot Interaction (HRI) [9]. The papers and discussion at that workshop make it evident that we should begin to see more and more research emerging at this intersection of fields.

In this paper, we summarize our own recent work on using augmented, mixed and virtual reality techniques to advance the state-of-the-art of robot-generated deixis, some of which was presented at the 2018 VAM-HRI workshop. In Sect. 2, we begin by providing a framework for categorizing robot-generated deixis in augmented and mixed-reality environments. In Sect. 3, we then discuss a novel method for enabling mixed reality deixis for armless robots. Finally, in Sect. 4 we present a novel method for robot teleoperation in Virtual Reality, and discuss how it could be used to trigger mixed-reality deictic gestures.

2 A Framework for Mixed-Reality Deictic Gesture

Augmented and mixed-reality technologies offer new opportunities for robots to communicate about the environments they share with human teammates. In previous work, we have presented a variety of work seeking to enable fluid natural language generation for robots operating in realistic human-robot interaction scenarios [10,11] (including work on referring expression generation [12,13], clarification request generation [14], and indirect speech act generation [15–17]). By augmenting their natural language references with visualizations that pick out their objects, locations, and people of interest within teammates' head-mounted displays, robots operating in such scenarios may facilitate conversational grounding [18,19] and shared mental modeling [20] with those human teammates in ways that were not previously possible.

While there has been some previous work on using visualizations as "gestures" within virtual or augmented environments [21] and video streams [22], as well as previous work on generating visualizations to accompany generated text [23–26], this metaphor of visualization-as-gesture has not yet been fully explored. This is doubly true for human-robot interaction scenarios, in which the use of augmented reality for human-robot communication is surprisingly underexplored. In fact, in their recent survey of augmented reality, Billinghurst et al. [27] cite intelligent systems, hybrid user interfaces, and collaborative systems as areas that have been under-attended-to in the AR community.

Most relevant to the current paper, Sibertsiva et al. [28] use augmented reality annotations to indicate different candidates referential hypotheses after receiving ambiguous natural language commands, and Green et al. [29] present a system that uses augmented reality to facilitate human-robot discussion of a plan prior to execution. There have also been several recent approaches to using augmented reality to non-verbally communicate robots' intentions [30–36] These approaches, however, have looked at visualization alone, outside the context of traditional robot gesture. We believe that, just as augmented and mixed reality open up new avenues for communication in human-robot interaction, human-robot interaction opens up new avenues for communication in augmented and mixed reality. Only in *mixed-reality human-robot interaction* may physical and virtual gestures be generated together or chosen between as part of a single process. In order to understand the different types of gestures that can be used in mixed-reality human-robot interaction, we have been developing a framework for analyzing such gestures along dimensions such as embodiment, cost, privacy, and legibility [37]. In this paper, we extend that framework to encompass new gesture categories and dimensions of analysis.

2.1 Conceptual Framework

In this section, we present a conceptual framework for describing mixed-reality deictic gestures. A robot operating within a pure-reality environment has access to but a single interface for generating gestures (its own body) and accordingly but a single perspective within which to generate them (its own)[1]. A robot operating within a mixed-reality environment, however, may leverage the hardware that enables such an environment, and the additional perspectives that come with those hardware elements. For robots operating within mixed-reality environments, we identify three unique hardware elements that can be used for deixis, each of which comes with its own perspective, and accordingly, their own class of deictic gestures.

First, robots may use their own bodies to perform the typical deictic gestures (such as pointing) available in pure reality. We categorize such gestures as *egocentric* (as shown in Fig. 1a), because they are generated from their own perspective. Second, robots operating in mixed-reality environments may be able to use of head-mounted displays worn by human teammates. We categorize such gestures as *allocentric* (as shown in Fig. 1b) because they are generated using only the perspective of the display's wearer. A robot, may, for example, "gesture" to an object by circling it within its teammate's display. Third, robots operating in mixed-reality environments may be able to use projectors to change how the world is perceived for all observers. We categorize such gestures as *perspective-free* (as shown in Fig. 1c) because they are not generated from the perspective of any one agent.

[1] Excepting, for the purposes of this paper, robots who are distributed across multiple sub-bodies in the environment [38].

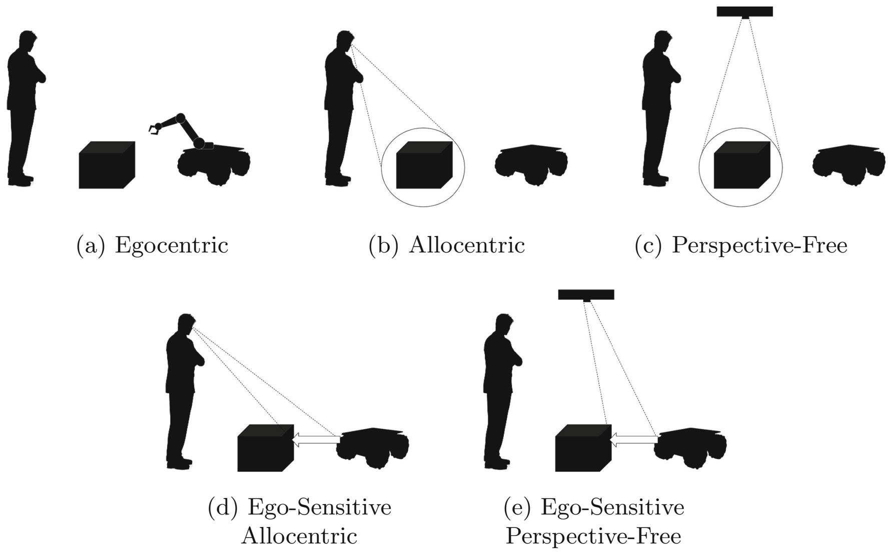


Fig. 1. Categories of mixed-reality deictic gestures

In addition, robots operating in mixed-reality environments may be able to perform multi-perspective gestures that use the aforementioned mixed-reality hardware in a way that connects back to the robot's perspectives. A robot may, for example, gesture to an object in its teammate's display, or using a projector, by drawing an arrow from itself to its target object, or by gesturing towards its target using a virtual appendage that only exists in virtuality. We call the former class *ego-sensitive allocentric gestures* and the latter class *ego-sensitive perspective-free gestures*.

Table 1. Analysis of mixed-reality deictic gestures

Category	HW	CG	Per	Emb	Cap	Pri (L)	Pri (G)	Cost (G)	Cost (M)	Leg (D)	Leg (S)
Ego	Rob	Yes	Rob	Yes	Yes	Low	High	High	Low	Low	Low
Allo	HMD	No	Hum	No	No	High	Low	Low	High	High	High
P-F	Pro	No	Env	No	No	Low	Low	Low	Low	High	High
ES Allo	HMD	Yes	Rob+Hum	Yes	No	High	Low	Low	High	TED	ED
ES P-F	Pro	Yes	Rob+Env	Yes	No	Low	Low	Low	Low	TED	ED

Dimensions: HW = Hardware; CG = Connection to Generator; Per = Perspective; Emb = Embodiment; Cap = Capability; Pri = Privacy (Local/Global); Cost; Leg = Legibility (Dynamic/Static).
Perspectives: Ego = Egocentric; Allo = Allocentric; P-F = Perspective-Free; ES = Ego-Sensitive.
Features: Rob = Robot; HMD = Head-Mounted Display; Pro = Projector; Hum = Human; Env = Environment; TED = Time and extent dependent; ED = Extent dependent.

2.2 Analysis of Mixed-Reality Deictic Gestures

Each of these gestural categories comes with its own unique properties. Here, we specifically examine six: perspective, embodiment, capability, privacy, cost, and legibility. These dimensions are summarized in Table 1.

The most salient dimensions that differentiate these categories of mixed-reality deictic gestures are the perspectives, embodiment, and capabilities they require. The perspectives required are clearly defined: egocentric gestures require access to the robot's perspective, allocentric gestures require access to the human interlocutor's perspective, and perspective-free gestures require access only to the greater environment's perspective. The ego-sensitive gestures connect their initial perspective with that of the robot. Those categories generated from or connected to the perspective of the robot notably require the robot to be embodied and co-present with their interlocutor; but only the egocentric category requires the robot's embodied form to be capable of movement.

The different hardware needs of these categories result in different levels of privacy. Here, we distinguish between *local* privacy and *global* privacy. We describe those categories that use a head-mounted display as affording high local privacy, as gestures are only visible to the human teammate with whom the robot is communicating. This dimension is particularly important for human-robot interaction scenarios involving both sensitive user populations (e.g., elder care or education) or in adversarial scenarios (e.g., competitive [39], police [40], campus safety [41], or military domains (as in DARPA's "Silent Talk" program) [42]). On the other hand, we describe egocentric gestures as having high *global* privacy, as, unlike with the other categories, information about gestural data need not be sent over a network, and thus may not be as vulnerable to hackers.

These categories of mixed-reality deictic gestures also come with different technical challenges, resulting in different computational costs. From the perspective of energy usage, egocentric gestures are expensive due to their physical component (a high *generation cost*). On the other hand, gestures that make use of a head-mounted display may be expensive to maintain due to registration challenges (a high *maintenance cost*).

Finally, these gestures differ with respect to legibility. In previous work, Dragan et al. [43] defined the notion of the legibility of an action, which describes the ease with which a human observer is able to determine the goal or purpose of an action as it is being carried out. In later work with Holladay et al. [5], Dragan then applies this notion to deictic gestures as well, analyzing the ability of the final gestural position to enable humans to pick out the target object. We believe, however, that this is really a distinct sense of legibility from Dragan's original formulation, and as such, we first refine this notion of legibility as applied to deictic gestures into two categories: we use *dynamic legibility* to refer to the degree to which a deictic gesture enables a human teammate to pick out the target object *as the action is unfolding* (in line with Dragan's original formulation), and *static legibility* to refer to the degree to which the final pose of a deictic gesture enables a human teammate to pick out the target object after the action is completed (in line with Holladay's formulation).

The gestural categories we describe differ with respect to both dynamic and static legibility. Allocentric and perspective-free gestures have high dynamic legibility (given that there is no dynamic dimension) and high static legibility (given that the target is uniquely picked out). Egocentric gestures have low dynamic legibility (relative to allocentric gestures) given that their target may not be clear at all as the action unfolds, and low static legibility, as the target may not be clear after the action is performed either, depending on distance to the target and density of distractors. The legibility of multi-perspective gestures depends on how exactly they are displayed. If they extend all the way to a target object, they may have high static legibility, whereas if they only point toward the target they will have low static legibility. Dynamic legibility depends both on this factor, as well as temporal extent. If a multi-perspective gesture unfolds over time, this may decrease the legibility (although it may better capture the user's attention).

2.3 Combination of Mixed-Reality Deictic Gestures

Finally, given these classes of mixed-reality deictic gestures, we can also reason about combinations of these gestures. Rather than explicitly discuss all 31 non-empty combinations of these five categories, we will briefly describe *how* the gestural categories combine. Simultaneous generation of gestures requiring different perspectives results in both perspectives being needed. The embodiment and capability requirements of simultaneous gestures combine disjunctively. The legibilities and costs of simultaneous gestures combine using a *max* operator, as the legibility of one gesture will excuse the illegibility of another, but the low cost of one gesture will not excuse the high cost of another. And the privacies of simultaneous gestures combine using a *min* operator, as the high privacy of one gesture does not excuse the low privacy of another.

3 Enabling Deictic Capabilities for Armless Robots Using Mixed-Reality Robotic Arms

In the previous section, we presented a framework for analyzing mixed-reality deictic gestures. Within this framework, the gestural categories that have received the least amount of previous attention are the ego-sensitive categories which connect the gesture-generating robot with the perspective of the human viewer or with the perspective of their environment. In this section, we present a novel approach to *ego-sensitive allocentric gesture*. Specifically, we propose to superimpose mixed-reality visualizations of robot arms onto otherwise armless robots, to allow them to gesture within their environment. This will allow an armless robot like a wheelchair or drone to gesture just as if it had a physical arm, even if mounting such an arm would not be mechanically possible or cost effective. Unlike purely allocentric gestures (e.g., circling an object in ones' field of view), this approach emphasizes the generator's embodiment, and as such, we would expect it to lead to increased perception of the robot's agency, increased likability of the robot, and promote positive team dynamics.

In this section we present the preliminary technical work necessary to enable such an approach. Specifically, we present a kinematic approach to perform this kind of mixed-reality deictic gesture. Compared to motion planning, a purely kinematic approach is more computationally efficient, a potential advantage for low-power embedded systems that we may wish to use for AR displays. The trade-off is that the kinematic approach is *incomplete*, so it may fail to find collision-free motions for some cluttered environments. However, collisions are not an impediment for virtual arms, thus mitigating the potential downside of purely kinematic motions.

Our approach applies dual-quaternion forward kinematics and Jacobian damped-least-squares inverse kinematics.

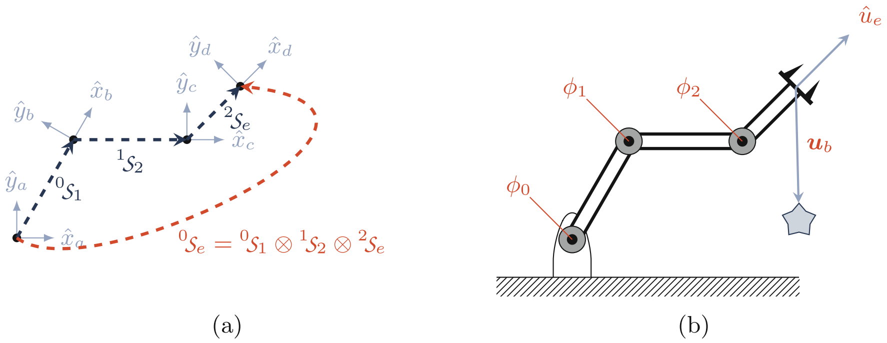


Fig. 2. Kinematic diagrams for diectic gestures. (a) the local coordinate frames ("frames") of a serial manipulator. (b) a schematic of a serial manipulator with vectors for pointing direction and the vector to a target object.

3.1 Kinematics

Forward Kinematics. We adopt the conventional model for serial robot manipulators of kinematic chains and trees [44–49]. Each local coordinate frame ("frame") of the robot has an associated label, and the frames are connected by Euclidean transformations (see Fig. 2a).

We represent Euclidean transformations with dual quaternions. Compared to matrix representations, dual quaternions offer computational advantages in efficiency, compactness, and numerical stability. A dual quaternion is a pair of quaternions: an *ordinary* part for rotation and *dual* part for translation. Notationally, we use a leading superscript to denote the parent's local coordinate frame p and trailing subscript to denote the child frame c. Given rotation unit quaternion $\hslash$ and translation vector $\mathbf{v}$ from p to c, the transformation dual quaternion ${}^p\mathcal{S}_c$ is:

$$\hbar = \left(h_x\hat{\boldsymbol{\imath}} + h_y\hat{\boldsymbol{\jmath}} + h_z\hat{\boldsymbol{k}} + h_w\right)$$
$$\mathbf{v} = \left(v_x\hat{\boldsymbol{\imath}} + v_y\hat{\boldsymbol{\jmath}} + v_z\hat{\boldsymbol{k}} + 0\right)$$
$${}^{p}\mathcal{S}_c = \left[\hbar + \left(\frac{1}{2}\mathbf{v} \otimes \hbar\right)\boldsymbol{\varepsilon}\right] \tag{1}$$

where $\hat{\boldsymbol{\imath}}$, $\hat{\boldsymbol{\jmath}}$, $\hat{\boldsymbol{k}}$ are the imaginary elements, with $\hat{\boldsymbol{\imath}}^2 = \hat{\boldsymbol{\jmath}}^2 = \hat{\boldsymbol{k}}^2 = \hat{\boldsymbol{\imath}}\hat{\boldsymbol{\jmath}}\hat{\boldsymbol{k}} = -1$, and $\boldsymbol{\varepsilon}$ is the dual element, with $\boldsymbol{\varepsilon}^2 = 0$ and $\boldsymbol{\varepsilon} \neq 0$.

Chaining transformations corresponds to multiplication of the transformation matrices or the dual quaternion. For a kinematic chain, we must match the child frame of predecessor to parent frame of successor transformations. The result is the transform from the parent of the initial to the child of the final transformation.

$${}^{a}\mathcal{S}_b \otimes {}^{b}\mathcal{S}_c = {}^{a}\mathcal{S}_c \tag{2}$$

We illustrate the kinematics computation for the simple serial manipulator in Fig. 2b. Note that the local frames and relative transforms of the robot in Fig. 2b correspond to those drawn in Fig. 2a.

The kinematic position of a robot is fully determined by its configuration $\boldsymbol{\phi}$, i.e, the vector of joint angles,

$$\boldsymbol{\phi} = [\phi_0, \phi_1, \ldots, \phi_n]^T \tag{3}$$

The relative frame at each joint i is a function of the corresponding configuration: ${}^{i-1}\mathcal{S}_i(\phi_i)$. The frame for the end-effector is the product of all frames in the chain

$${}^{0}\mathcal{S}_e(\boldsymbol{\phi}) = \left({}^{0}\mathcal{S}_1(\phi_0)\right) \otimes \left({}^{1}\mathcal{S}_2(\phi_1)\right) \otimes \ldots \otimes \left({}^{n-1}\mathcal{S}_n(\phi_{n-1})\right) \otimes \left({}^{n}\mathcal{S}_e(\phi_n)\right) \tag{4}$$

Cartesian Control. We compute the least-squares solution for Cartesian motion using a singularity-robust Jacobian pseudoinverse:

$$\dot{\mathbf{x}} = \mathbf{J}\dot{\boldsymbol{\phi}} \quad \rightsquigarrow \quad \dot{\boldsymbol{\phi}} = \mathbf{J}^{+}\dot{\mathbf{x}} \tag{5}$$

$$\mathbf{J}^{+} = \sum_{i=0}^{\min(m,n)} \frac{s_i}{\max({s_i}^2, {s_{\min}}^2)} \mathbf{v_i}\mathbf{u_i}^T \tag{6}$$

where $\dot{\mathbf{x}} = [\omega, \dot{v}]$ is the vector of rotational velocity ω and translational velocity $\dot{v}$, $\mathbf{J} = \mathbf{USV}^T$ is the singular value decomposition[2] of Jacobian J, and $s_{\min}$ is a selected constant for the minimum acceptable singular value.

We determine Cartesian velocity $\dot{\mathbf{x}}$ with a proportional gain on position error, computed as the velocity to reach the desired target in unit time, decoupling rotational ω and translational $\dot{v}$ parts to achieve straight-line translations:

$$\hbar_1 = \exp\left(\frac{\omega \Delta t}{2}\right) \otimes \hbar_0 \quad \rightsquigarrow \quad -\omega \Delta t = 2\ln\left(\hbar_0 \otimes \hbar_1^*\right) \tag{7}$$

[2] The SVD, while more expensive to compute, provides better accuracy and numerical stability for Cartesian control than the LU decomposition.

$$v_1 = \dot{x}\Delta t + v_0 \quad \rightsquigarrow \quad -\dot{x}\Delta t = v_0 - v_1 \tag{8}$$

In combination, we compute the reference joint velocity as:

$$\dot{\phi} = \mathbf{J}^{+}\left(-k\begin{bmatrix} 2\ln\left({}^{0}\hbar_e \otimes \left({}^{0}\hbar_r\right)^{*}\right) \\ {}^{0}v_e - {}^{0}v_r \end{bmatrix}\right) \tag{9}$$

where e is the actual end-effector frame and r is the desired or reference frame.

3.2 Design Patterns

While the method above provides a general approach to enabling mixed-reality deictic gestures, there are a variety of different possible forms of deictic gestures that might be generated using that approach. In this section, we propose three candidate gesture designs enabled by the proposed approach: *Fixed Translation*, *Reaching*, and *Floating*.

Fixed Translation. The first proposed design, *Fixed Translation*, is the most straightforward manifestation of the proposed approach. In this design, the visualized arm rotates in place to point to the desired target. To enable this design, we must find a target orientation ${}^{0}\hbar_r$ for the end-effector. We find the relative rotation ${}^{e}\hbar_r$ between the current end-effector frame and pointing direction towards the target based on the end-effector's pointing direction vector and the vector from the end-effector to the target (see Fig. 2b).

First, we find the end-effector's global pointing vector $\hat{u}_e$ by rotating the local pointing direction $\hat{a}_e$.

$$\hat{u}_e = {}^{0}\hbar_e \otimes \hat{a}_e \otimes \left({}^{0}\hbar_e\right)^{*} \tag{10}$$

Then, we find the vector from the end-effector to the target by subtracting the end-effector translation ${}^{0}v_e$ from the target translation ${}^{0}v_b$ and normalizing to a unit vector.

$$\begin{aligned} {}^{0}v_e &= 2\ {}^{0}d_e \otimes \left({}^{0}\hbar_e\right)^{*} \\ \hat{u}_b &= \frac{{}^{0}v_b - {}^{0}v_e}{\|{}^{0}v_b - {}^{0}v_e\|} \end{aligned} \tag{11}$$

Next, we compute the relative rotation between the two vectors $\hat{u}_e$ and $\hat{u}_b$ using the dot product to find the angle θ and cross product to find the axis $\hat{a}$,

$$\theta = \cos^{-1}\left(\hat{u}_e \bullet \hat{u}_b\right) \tag{12}$$

$$\hat{a} = \frac{\hat{u}_e \times \hat{u}_b}{\sin\theta} \tag{13}$$

The axis $\hat{a}$ and angle θ then give us the rotation unit quaternion ${}^{e}\hbar_b$:

$${}^{e}\hbar_r = \left(\hat{a}\sin\frac{\theta}{2} + \cos\frac{\theta}{2}\right) \tag{14}$$

Note that a direct conversion of the vectors to the rotation unit quaternion avoids the need for explicit evaluation of transcendental functions.

Now, we compute the global reference frame for the end-effector using

$$
\begin{aligned}
{}^{e}\mathcal{S}_r &= \left({}^{e}\hbar_b + 0\boldsymbol{\varepsilon}\right) \\
{}^{0}\mathcal{S}_r &= {}^{0}\mathcal{S}_e \otimes {}^{e}\mathcal{S}_b
\end{aligned} \tag{15}
$$

Combining, (15) and (9), we compute the joint velocities $\dot{\boldsymbol{\phi}}$ for the robot arm.

Reaching. Our second proposed design, *Reaching*, stretches the arm out towards the target, increasing gesture legibility in a way that would not be feasible with a physical arm. To enable this design, we compute the instantaneous desired orientation as in the fixed translation case, but now set the desired translation to the target object's translation ${}^{0}v_b$.

$$
\begin{aligned}
{}^{0}\hbar_r &= {}^{0}\hbar_e \otimes {}^{e}\hbar_b \\
{}^{0}\mathcal{S}_r &= \left({}^{0}\hbar_r + \left(\frac{1}{2}{}^{0}\hbar_r \otimes {}^{0}v_b\right)\boldsymbol{\varepsilon}\right)
\end{aligned} \tag{16}
$$

Then we combine, (16) and (9) to compute the joint velocities $\dot{\boldsymbol{\phi}}$ for the robot arm.

Floating Translation. Diectic information is conveyed primarily by the orientation of the end-effector rather than its translation. Thus, in our final design, *Floating Translation*, we consider a case where the translation can freely float, allowing the arm to point with more natural-looking configurations. First, we remove the translational component from the control law. Second, we center all joints within the Jacobian null space, so centering does not impact end-effector velocity. We update the workspace control law with a weighting matrix and null space projection term:

$$
\dot{\boldsymbol{\phi}} = \mathbf{J}^{+}\mathbf{W}\dot{\mathbf{x}} + \left(\mathbf{I} - \mathbf{J}^{+}\mathbf{J}\right)\dot{\boldsymbol{\phi}}_N \tag{17}
$$

The weighting matrix $\mathbf{W}$ removes the translational component from the Jacobian $\mathbf{J}$, so only rotational error contributes to the joint velocity $\dot{\boldsymbol{\phi}}$. Structurally, $\mathbf{J}^{+}$ consists of rotational block $\mathbf{j}_{\omega}^{+}$ and translational block $\mathbf{j}_{\dot{v}}^{+}$. We construct $\mathbf{W}$ to remove $\mathbf{j}_{\dot{v}}^{+}$.

$$
\begin{aligned}
\mathbf{J}^{+} &= \left[\mathbf{j}_{\omega}^{+} \mid \mathbf{j}_{\dot{v}}^{+}\right] \\
\mathbf{W} &= \begin{bmatrix}\mathbf{I}_{3\times n} \\ \mathbf{0}_{3\times n}\end{bmatrix} \\
\mathbf{J}^{+}\mathbf{W} &= \left[\mathbf{j}_{\omega}^{+} \mid \mathbf{0}\right]
\end{aligned} \tag{18}
$$

where n is the length of $\boldsymbol{\phi}$, or equivalently the number of rows in $\mathbf{J}^{+}$.

We use the null space projection to move all joints towards their center configuration, without impacting on end-effector pose:

$$\dot{\boldsymbol{\phi}} = \boldsymbol{\phi}_c - \boldsymbol{\phi}_a \tag{19}$$

where $\boldsymbol{\phi}_c$ is the center configuration and $\boldsymbol{\phi}_a$ is the actual configuration.

The combined workspace control law is

$$\dot{\boldsymbol{\phi}} = \left(\mathbf{J}^{+}\right) \begin{bmatrix} \mathbf{I}_{3\times n} \\ \mathbf{0}_{3\times n} \end{bmatrix} \dot{\mathbf{x}} + \left(\mathbf{I} - \mathbf{J}^{+}\mathbf{J}\right)\left(\boldsymbol{\phi}_c - \boldsymbol{\phi}_a\right) \tag{20}$$

In this paper, we have proposed a new form of mixed-reality deictic gesture, and proposed a space of candidate designs for manifesting such gestures. In current and future work, we will implement all three designs using the Microsoft Hololens, and evaluate their performance with respect to both each other, and to the other categories of gesture we have described. In the next section, we turn to methods by which such gestures might be generated by human teleoperators during human-subject experiments.

4 An Interface for Virtual Reality Teleoperation

In the previous sections, we presented a framework for mixed-reality deixis, and a novel form of mixed-reality deictic gesture. But a question remains as to how robots might decide to generate such gestures. While in future work our interests lie in computational approaches for allowing robots to decide for themselves when and how to generate such gestures, in this work we first examine how humans might trigger such gestures, and how novel *virtual reality* technologies might facilitate this process.

Specifically, we examine the use of virtual reality and gesture recognition technologies may be used to control gesture-capable robots used by Human-Robot Interaction (HRI) researchers during human-subject experiments [50]. Manual control of language- and gesture-capable robots is crucial for HRI researchers seeking to evaluate human perceptions of potential autonomous capabilities which either do not yet exist, or are not yet robust enough to work consistently and predictably, as in the *Wizard of Oz* (WoZ) experimental paradigm [51]. For the purposes of such experiments, manual control of dialogue and gestural capabilities is particularly challenging [52]. Not only is it repetitive and time consuming to design WoZ interfaces for such capabilities, but such interfaces are not always effective, as the time necessary for an experimenter to decide to issue a command, click the appropriate button, and have that command take effect on the robot is typically too long to facilitate natural interaction.

What is more, such interfaces typically require experimenters to switch back and forth between monitoring a camera stream depicting the robot's environment and consulting their control interface: a pattern that can decrease robots' situational awareness and harm experiment effectiveness [53]. This is particularly true when the camera stream depicts the robot's environment from a

(a) The teleoperation interface, and its integration with the Softbank Pepper robot.

(b) Simulated stereoscopic display displayed to the user through the VR headset.

(c) View of gesture tracking data in the Leap Motion SDK Diagnostic Visualizer.

Fig. 3. Multiple views of integrated system

third-person perspective, which can lead to serious performance challenges [54]. While some recent approaches have introduced the use of augmented reality for safely teleoperating co-present robots [55,62], robots are not typically co-present with teleoperators during tightly controlled WoZ experiments. For such applications, Virtual Reality (VR) teleoperation provides one possible solution. VR is also beneficial as immersion in the robot's perspective improves depth perception and enhances visual feedback, resulting in an overall more immersive experience [56]. On the other hand, immersive first-person teleoperation comes with its own concerns. Recent researchers have noted safety concerns, as a sufficiently constrained robot perspective may limit the teleoperator's situational awareness [57]. What is more, VR teleoperation in particular raises challenges as the teleoperator may no longer be able see their teleoperation interface.

4.1 Previous Work

There have been a large number of approaches to robot teleoperation through virtual reality, even within only the past year. First, there has been some work on robot teleoperation using touchscreens displaying first- or third-person views of the robot's environment [63,64]. There have been a number of approaches enabling first-person robot teleoperation using virtual reality displays, using a variety of different control modalities, including joysticks [65], VR hand controllers [66–71], gloves [72,73], and full-torso exo-suits [59]. There has been less work enabling hands-free teleoperation, with the closest previous work we are aware of being Miner and Stansfield's approach, which allowed gesture-based control in *simulated, third-person* virtual reality. The only approaches we are aware of enabling first-person hands-free control are our own approach (discussed in the next section), and the Kinect-based approach of Sanket et al., which was presented at the same workshop as our own work [70].

4.2 Integrated Approach

In our recent work [50], we have proposed a novel teleoperation interface which provides hands-free WoZ control of a robot while providing the teleoperator with

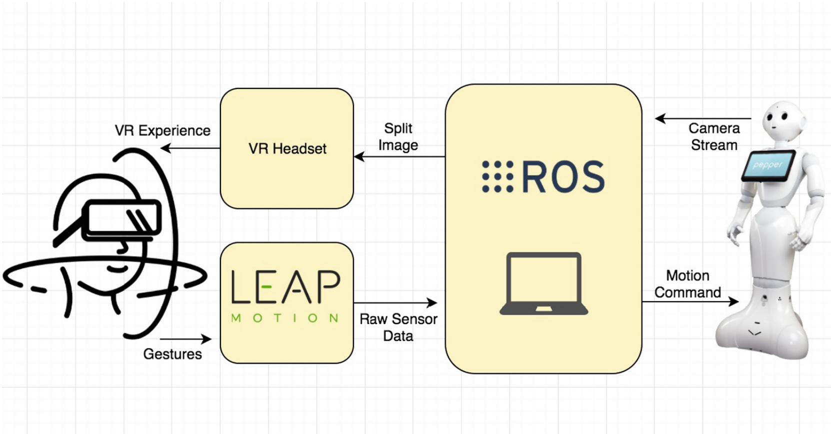


Fig. 4. Architecture diagram: The user interacts directly with a VR headset (e.g., Google Cardboard) and a Leap Motion gesture sensor. These devices send data to and receive data from a humanoid robot (e.g., the Softbank Pepper) using an instance of the ROS architecture whose Master node is run on a standard Linux laptop.

an immersive VR experience from the robot's point of view. This interface integrates a VR headset, interfaced directly with the robot's camera to allowing the experimenter to see exactly what the robot sees (Fig. 3b), with a Leap Motion Controller. Translating traditional joystick or gamepad control to robotic arm motions can be challenging, but the Leap Motion Controller can simplify this process by allowing the user to replicate the gesture he/she desires of the robot, making it a powerful hands-free teleoperation device [58]. There has been work on using the Leap Motion for teleoperation *outside* the context of virtual reality [74–76] but to the best of our knowledge our approach is the first to pair it with an immersive virtual-reality display. In our approach, we use the Leap Motion sensor to capture the experimenter's gestures, and then generate analogous gestures on the robot in real time. Specifically, we first extract hand position and orientation data from raw Leap Motion data. Figure 3c shows the visualization of the tracking data produced by the Leap Motion. Each arrow represents a finger, and each trail represents the corresponding movement of that finger. Changes in this position and orientation data is used to trigger changes in the robot's gestures according to the following equations:

$$robotGesturePitch = \begin{cases} low & \tau_{p_1} < humanGesturePitch < \tau_{p_2} \\ high & \tau_{p_2} < humanGesturePitch < \tau_{p_3} \end{cases}$$

$$robotGestureRoll = \begin{cases} low & \tau_{r_1} < humanGestureRoll < \tau_{r_2} \\ high & \tau_{r_2} < humanGestureRoll < \tau_{r_3} \end{cases}$$

Here, parameters $\tau_{p_1} < \tau_{p_2} < \tau_{p_3}$ and $\tau_{r_1} < \tau_{r_2} < \tau_{r_3}$ are manually defined pitch and raw thresholds. While in this work our initial prototype makes use of these simple inequalities, in future work we aim to examine more sophisticated geometric and approximate methods for precisely mapping human gestures to robot gestures, with the aim of enabling a level of control currently seen in suit-based teleoperation systems [59].

All components of the proposed interface are integrated using the Robot Operating System (ROS) [60]. As shown in Fig. 4, the Leap Motion publishes raw sensor data, which is converted into motion commands. These motion commands are then sent to the robot[3]. Similarly, camera data is published by the robot, to a topic subscribed to by the Android VR app which displays it in the VR headset[4].

5 Conclusion

Virtual, augmented, and mixed reality stand to enable – and are already enabling – promising new paradigms for human-robot interaction. In this work, we summarized our own recent work in all three of these areas. We see a long, bright avenue for future work in this area for years to come. In our own future work, we plan to focus on exploring the space of different designs for mixed-reality deictic gesture, and integrating these approaches with our existing body of work on natural language generation, thus enabling exciting new ways for robots to express themselves.

References

1. McNeill, D.: Hand and Mind: What Gestures Reveal About Thought. University of Chicago Press, Chicago (1992)
2. Fillmore, C.J.: Towards a descriptive framework for spatial deixis. In: Speech, Place and Action: Studies in Deixis and Related Topics, pp. 31–59 (1982)
3. Salem, M., Kopp, S., Wachsmuth, I., Rohlfing, K., Joublin, F.: Generation and evaluation of communicative robot gesture. Int. J. Soc. Robot. **4**(2), 201–217 (2012)
4. Huang, C.M., Mutlu, B.: Modeling and evaluating narrative gestures for humanlike robots. In: Robotics: Science and Systems, pp. 57–64 (2013)
5. Holladay, R.M., Dragan, A.D., Srinivasa, S.S.: Legible robot pointing. In: 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 217–223. IEEE (2014)
6. Sauppé, A., Mutlu, B.: Robot deictics: how gesture and context shape referential communication. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, pp. 342–349. ACM (2014)
7. Gulzar, K., Kyrki, V.: See what i mean-probabilistic optimization of robot pointing gestures. In: Proceedings of the Fifteenth IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 953–958. IEEE (2015)

[3] While in this work we use the Softbank Pepper robot, our general framework is not necessarily specific to this particular robot.

[4] In this work, we use a single camera, as Pepper has a single RGB camera rather than stereo cameras. In the future, we hope to use stereoscopic vision as input for a more immersive VR experience. In addition, images could be just as easily streamed to this app from a ROS simulator (e.g., Gazebo [61]) or from some other source.

8. Admoni, H., Weng, T., Scassellati, B.: Modeling communicative behaviors for object references in human-robot interaction. In: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pp. 3352–3359. IEEE (2016)
9. Williams, T., Szafir, D., Chakraborti, T., Ben Amor, H.: Virtual, augmented, and mixed reality for human-robot interaction. In: Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 403–404. ACM (2018)
10. Williams, T.: A consultant framework for natural language processing in integrated robot architectures. IEEE Intell. Inform. Bull. **18**(1), 10–14 (2017)
11. Williams, T.: Situated natural language interaction in uncertain and open worlds. Ph.D. thesis, Tufts University (2017)
12. Kunze, L., Williams, T., Hawes, N., Scheutz, M.: Spatial referring expression generation for HRI: Algorithms and evaluation framework. In: AAAI Fall Symposium on AI and HRI (2017)
13. Williams, T., Scheutz, M.: Referring expression generation under uncertainty: algorithm and evaluation framework. In: Proceedings of the 10th International Conference on Natural Language Generation (2017)
14. Williams, T., Scheutz, M.: Resolution of referential ambiguity in human-robot dialogue using dempster-shafer theoretic pragmatics. In: Proceedings of Robotics: Science and Systems (2017)
15. Briggs, G., Williams, T., Scheutz, M.: Enabling robots to understand indirect speech acts in task-based interactions. J. Hum.-Robot Interact. **26**, 64–94 (2017)
16. Williams, T., Briggs, G., Oosterveld, B., Scheutz, M.: Going beyond literal command-based instructions: extending robotic natural language interaction capabilities. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence (2015)
17. Williams, T., Thames, D., Novakoff, J., Scheutz, M.: Thank you for sharing that interesting fact!: effects of capability and context on indirect speech act use in task-based human-robot dialogue. In: Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (2018)
18. Fussell, S.R., Setlock, L.D., Kraut, R.E.: Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 513–520. ACM (2003)
19. Kraut, R.E., Fussell, S.R., Siegel, J.: Visual information as a conversational resource in collaborative physical tasks. Hum.-Comput. Interact. **18**(1), 13–49 (2003)
20. Datcu, D., Cidota, M., Lukosch, H., Lukosch, S.: On the usability of augmented reality for information exchange in teams from the security domain. In: 2014 IEEE Joint Intelligence and Security Informatics Conference (JISIC), pp. 160–167. IEEE (2014)
21. White, S., Lister, L., Feiner, S.: Visual hints for tangible gestures in augmented reality. In: 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 47–50. IEEE (2007)
22. Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., Kramer, A.D.: Gestures over video streams to support remote collaboration on physical tasks. Hum.-Comput. Interact. **19**(3), 273–309 (2004)
23. Wahlster, W., André, E., Graf, W., Rist, T.: Designing illustrated texts: how language production is influenced by graphics generation. In: Proceedings of the Fifth Conference on European Chapter of the Association for Computational Linguistics, pp. 8–14. Association for Computational Linguistics (1991)

24. Wazinski, P.: Generating spatial descriptions for cross-modal references. In: Proceedings of the Third Conference on Applied Natural Language Processing, pp. 56–63. Association for Computational Linguistics (1992)
25. Green, S.A., Billinghurst, M., Chen, X., Chase, J.G.: Human robot collaboration: an augmented reality approacha literature review and analysis. In: Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 117–126. American Society of Mechanical Engineers (2007)
26. Green, S., Billinghurst, M., Chen, X., Chase, G.: Human-robot collaboration: a literature review and augmented reality approach in design. Int. J. Adv. Robot. Syst. **5**, 1 (2008)
27. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. Found. Trends Hum.-Comput. Interact. **8**(2–3), 73–272 (2015)
28. Sibirtseva, E., Kontogiorgos, D., Nykvist, O., Karaoguz, H., Leite, I., Gustafson, J., Kragic, D.: A comparison of visualisation methods for disambiguating verbal requests in human-robot interaction. arXiv preprint arXiv:1801.08760 (2018)
29. Green, S.A., Chase, J.G., Chen, X., Billinghurst, M.: Evaluating the augmented reality human-robot collaboration system. Int. J. Intell. Syst. Technol. Appl. **8**(1–4), 130–143 (2009)
30. Andersen, R.S., Madsen, O., Moeslund, T.B., Amor, H.B.: Projecting robot intentions into human environments. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 294–301. IEEE (2016)
31. Chadalavada, R.T., Andreasson, H., Krug, R., Lilienthal, A.J.: That's on my mind! robot to human intention communication through on-board projection on shared floor space. In: 2015 European Conference on Mobile Robots (ECMR), pp. 1–6. IEEE (2015)
32. Frank, J.A., Moorhead, M., Kapila, V.: Mobile mixed-reality interfaces that enhance human-robot interaction in shared spaces. Front. Robot. AI **4**, 20 (2017)
33. Ganesan, R.K.: Mediating human-robot collaboration through mixed reality cues. Master's thesis, Arizona State University (2017)
34. Katzakis, N., Steinicke, F.: Excuse me! perception of abrupt direction changes using body cues and paths on mixed reality avatars. arXiv preprint arXiv:1801.05085 (2018)
35. Rosen, E., Whitney, D., Phillips, E., Chien, G., Tompkin, J., Konidaris, G., Tellex, S.: Communicating robot arm motion intent through mixed reality head-mounted displays. arXiv preprint arXiv:1708.03655 (2017)
36. Walker, M., Hedayati, H., Lee, J., Szafir, D.: Communicating robot motion intent with augmented reality. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 316–324. ACM (2018)
37. Williams, T.: A framework for robot-generated mixed-reality deixis. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
38. Oosterveld, B., Brusatin, L., Scheutz, M.: Two bots, one brain: component sharing in cognitive robotic architectures. In: Proceedings of 12th ACM/IEEE International Conference on Human-Robot Interaction Video Contest (2017)
39. Correia, F., Alves-Oliveira, P., Maia, N., Ribeiro, T., Petisca, S., Melo, F.S., Paiva, A.: Just follow the suit! trust in human-robot interactions during card game playing. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 507–512. IEEE (2016)

40. Bethel, C.L., Carruth, D., Garrison, T.: Discoveries from integrating robots into swat team training exercises. In: 2012 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), pp. 1–8. IEEE (2012)
41. Goldfine, S.: Assessing the prospects of security robots, October 2017
42. Kotchetkov, I.S., Hwang, B.Y., Appelboom, G., Kellner, C.P., Connolly Jr., E.S.: Brain-computer interfaces: military, neurosurgical, and ethical perspective. Neurosurg. Focus **28**(5), E25 (2010)
43. Dragan, A.D., Lee, K.C., Srinivasa, S.S.: Legibility and predictability of robot motion. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 301–308. IEEE (2013)
44. Şucan, I.A., Chitta, S.: MoveIt! (2015). http://moveit.ros.org
45. Dantam, N.T., Chaudhuri, S., Kavraki, L.E.: The task motion kit. Robot. Autom. Mag. (2018, accepted)
46. Diankov, R.: Automated construction of robotic manipulation programs. Ph.D. thesis, Carnegie Mellon University, Robotics Institute, August 2010
47. Hartenberg, R.S., Denavit, J.: Kinematic Synthesis of Linkages. McGraw-Hill, New York (1964)
48. Smits, R., Bruyninckx, H., Aertbeliën, E.: KDL: Kinematics and dynamics library (2011). http://www.orocos.org/kdl
49. Willow Garage: URDF XML (2013). http://wiki.ros.org/urdf/XML
50. Tran, N., Rands, J., Williams, T.: A hands-free virtual-reality teleoperation interface for wizard-of-oz control. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
51. Riek, L.D.: Wizard of oz studies in HRI: a systematic review and new reporting guidelines. J. Hum.-Robot Interact. **1**(1), 119–136 (2012)
52. Bonial, C., Marge, M., Foots, A., Gervits, F., Hayes, C.J., Henry, C., Hill, S.G., Leuski, A., Lukin, S.M., Moolchandani, P., et al.: Laying down the yellow brick road: development of a wizard-of-oz interface for collecting human-robot dialogue. arXiv preprint arXiv:1710.06406 (2017)
53. Chen, J.Y.C., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **37**(6), 1231–1245 (2007)
54. Nawab, A., Chintamani, K., Ellis, D., Auner, G., Pandya, A.: Joystick mapped augmented reality cues for end-effector controlled tele-operated robots. In: 2007 IEEE Virtual Reality Conference, pp. 263–266, March 2007
55. Hedayati, H., Walker, M., Szafir, D.: Improving collocated robot teleoperation with augmented reality. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 78–86. ACM (2018)
56. Miner, N.E., Stansfield, S.A.: An interactive virtual reality simulation system for robot control and operator training. In: Proceedings of the 1994 IEEE International Conference on Robotics and Automation, vol. 2, pp. 1428–1435, May 1994
57. Rakita, D., Mutlu, B., Gleicher, M.: An autonomous dynamic camera method for effective remote teleoperation. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, pp. 325–333. ACM (2018)
58. Pititeeraphab, Y., Choitkunnan, P., Thongpance, N., Kullathum, K., Pintavirooj, C.: Robot-arm control system using leap motion controller. In: 2016 International Conference on Biomedical Engineering (BME-HUST), pp. 109–112, October 2016
59. Bennett, M., Williams, T., Thames, D., Scheutz, M.: Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2017)

60. Quigley, M., Faust, J., Foote, T., Leibs, J.: ROS: an open-source robot operating system. In: ICRA Workshop on Open Source Software (2009)
61. Koenig, N., Howard, A.: Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2149–2154. IEEE (2004)
62. Gong, L., Gong, C., Ma, Z., Zhao, L., Wang, Z., Li, X., Jing, X., Yang, H., Liu, C.: Real-time human-in-the-loop remote control for a life-size traffic police robot with multiple augmented reality aided display terminals. In: 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 420–425, August 2017
63. Quintero, C.P., Ramirez, O.A., Jägersand, M.: VIBI: assistive vision-based interface for robot manipulation. In: ICRA, pp. 4458–4463 (2015)
64. Hashimoto, S., Ishida, A., Inami, M., Igarashi, T.: TouchMe: an augmented reality based remote robot manipulation. In: The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011, vol. 2 (2011)
65. Pereira, A., Carter, E.J., Leite, I., Mars, J., Lehman, J.F.: Augmented reality dialog interface for multimodal teleoperation. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (2017)
66. Haring, K.S., Finomore, V., Muramato, D., Tenhundfeld, N.L., Redd, M., Wen, J., Tidball, B.: Analysis of using virtual reality (VR) for command and control applications of multi-robot systems. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
67. Lipton, J.I., Fay, A.J., Rus, D.: Baxter's homunculus: virtual reality spaces for teleoperation in manufacturing. IEEE Robot. Autom. Lett. **3**(1), 179–186 (2018)
68. Oh, Y., Parasuraman, R., McGraw, T., Min, B.C.: 360 VR based robot teleoperation interface for virtual tour. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
69. Rosen, E., Whitney, D., Phillips, E., Ullman, D., Tellex, S.: Testing robot teleoperation using a virtual reality interface with ROS reality. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
70. Gaurav, S., Al-Qurashi, Z., Barapatre, A., Ziebart, B.: Enabling effective robotic teleoperation using virtual reality and correspondence learning via neural network. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
71. Whitney, D., Rosen, E., Phillips, E., Konidaris, G., Tellex, S.: Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality. In: Proceedings of the International Symposium on Robotics Research (2017)
72. Allspaw, J., Roche, J., Lemiesz, N., Yannuzzi, M., Yanco, H.A.: Remotely teleoperating a humanoid robot to perform fine motor tasks with virtual reality-. In: Proceedings of the 2018 Conference on Waste Management (2018)
73. Allspaw, J., Roche, J., Norton, A., Yanco, H.A.: Remotely teleoperating a humanoid robot to perform fine motor tasks with virtual reality-. In: Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI) (2018)
74. Bassily, D., Georgoulas, C., Guettler, J., Linner, T., Bock, T.: Intuitive and adaptive robotic arm manipulation using the leap motion controller. In: 41st International Symposium on Robotics ISR/Robotik 2014, pp. 1–7, June 2014

75. Lin, Y., Song, S., Meng, M.Q.H.: The implementation of augmented reality in a robotic teleoperation system. In: IEEE International Conference on Real-Time Computing and Robotics (RCAR), pp. 134–139. IEEE (2016)
76. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. Sensors **13**(5), 6380–6393 (2013)

Embodiment, Communication and Collaboration in VAMR

Is This Person Real? Avatar Stylization and Its Influence on Human Perception in a Counseling Training Environment

Jennie Ablanedo[1(✉)], Elaine Fairchild[2], Tami Griffith[3], and Christopher Rodeheffer[2]

[1] Institute for Simulation and Training (IST), University of Central Florida (UCF), Orlando, USA
jablaned@ist.ucf.edu
[2] Defense Equal Opportunity Management Institute (DEOMI), J-9 Research Directorate, Patrick Air Force Base, USA
{elaine.fairchild.ctr, christopher.rodeheffer}@us.af.mil
[3] Army Research Laboratory - Simulation and Training Technology Center, Orlando, USA
tamara.s.griffith.civ@mail.mil

Abstract. This paper describes a pilot study planned by the Defense Equal Opportunity Employment Management Institute (DEOMI). By leveraging previous work in maturing a low-cost real-time puppeted character in a virtual environment, the team is seeking to explore the role stylization plays in how participants perceive emotions and connect with an avatar emotionally within a training atmosphere. The paper also describes future work in exploring how biases might be exposed when interacting with puppeted virtual characters.

Keywords: Virtual training · Avatar puppeteering · Virtual humans
Real-time motion capture · Stylization · Culture bias

1 Introduction

1.1 Background

Current commercial games, such as Assassin's Creed, ® exemplify the incredible level of realism in virtual environments and the characters residing within them. However, that realism fades during in-depth interactions with characters employing artificial intelligence. Real-time puppeteering is one strategy

Fig. 1. Characters from Assassin's Creed® Origins: The Hidden Ones [25]

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 279–289, 2018.
https://doi.org/10.1007/978-3-319-91581-4_20

to increase the realism of those interactions within virtual environments. Puppeteering is a technique that allows the movements, expressions, and actions of a virtual avatar to be fully controlled by an individual through a motion capture system [1] (Fig. 1).

1.2 Avatar Realism

An avatar is a virtual representation of a human being, which is fully controlled by an actual human [2]. The aesthetics or visual representation of an avatar can vary based upon the artist depiction of their creation (see Stylization).

A study conducted by Trinity College Dublin [3] explored the appearance of realism and animation of an individual's own-virtual face creating character appeal. The findings of this study concluded that whether an avatar is realistic or displayed abstract cartoon-like features, the participant was not affected by the animation realism [3]. Realistic faces are equally appealing to cartoon faces in both animation realism conditions. However, when looking at an emphasis of perceived emotions, a study [4] analyzing different stylization techniques concluded that different avatar stylizations can affect the expression appearance and human response towards the perception of that Avatar.

The Threshold Model of Social Influence [5], states that the social influence of a real person is high, however, the influence of an Artificial Intelligence (AI) character depends on the character's realism. Furthering this research, Von der Putten [2] found that participants experienced greater negative feelings towards artificial agents than avatars who displayed more realistic features, in addition to a feeling of greater social presence. Their findings essentially state that avatars who aesthetically display more realistic features and behaviors will create a greater social awareness and perception of the virtual character.

By leveraging the work we have done utilizing avatar puppeteering [1], we are seeking to expand upon the listed study findings to see if the stylization produces an effect of how participants perceive emotions or an emotional connection with an avatar controlled by a puppeteer within a training atmosphere. We are also interested to see if the look of an avatar controlled by a puppeteer changes the perceived virtual presence and what biases may be exposed.

1.3 Motion Capture Technology

Motion capture systems are frequently used in the entertainment industry in the creation of animations for seamless, elegant, and fluid appearance of characters (See Fig. 2) [6]. The initial motion capture sequences are polished by animators to ensure movements appear clean and natural [7]. The use of this technology in real-time, without the clean-up sequence, is still in its infancy. Participants interacting with a character controlled by a puppeteer see the character display gestures and facial expressions as if they are speaking to someone in real life. The individual controlling the character can use a web camera to view the person with whom they are interacting. This can change the participant's experience from a witness to someone who is wrapped into the experience. By interacting with an avatar played by a human in real time, it may create a more personal and immersive emotional experience.

Fig. 2. War for the planet of the apes motion capture – twentieth century fox [26]

1.4 Puppeteering

The use of technology-based communication can pose unique challenges [8–10]. Specifically, the use of technology can lead to errors in interpretation and decreases in performance [8, 9]. One area that can be especially challenging is the interpretation of human expressions. The human face has forty-six unique action units that are capable of making more than ten thousand unique configurations and these configurations allow for greater understanding of the communicated message [11]. For example, a close friend may use a subtle eyebrow raise to accompany a statement. The statement alone might be considered offensive, but in conjunction with the eyebrow raise, it would be perceived to be a pleasant and playful jest. In addition to facial expressions, changes in voice tone can also change the context of a statement [11, 12].

The realism of interpersonal interactions in virtual training domains can be an important factor in influencing decision making in a learning population [13]. Some training tasks require very detailed realism, including the ability to recognize small-motor movements, eye movements, voice intonation and the gesticulations.

The interaction between automated virtual agents does not yet approach human-to-human experiences. This drives our team's interest in increasing the realism of human interaction in virtual environments by using real-time puppeteering. The focus is on merging state-of-the-art commercial technologies to support human-dimension training for force effectiveness. The working prototype provides a platform for feedback from end-users and informs the requirements and procurement communities. This capability is possible at a low cost, though the technology is still in its infancy.

To date, research on real-time puppeteering is limited. Even more limited is research on the impacts of puppeteering in a training environment. This is problematic as more organizations are relying on technology for training [14]. The goal of this research task is to explore the training implications of interacting with a character controlled by a puppeteer training certain human interactions in virtual environments. This paper examines: the areas of realism that might provide the greatest payoff per investment; the impact of stylization on perceptions of avatar realism and the overall effect of these perceptions on important organizational outcomes in regards to a training engagement and training performance; what expressions immerse the learner enough to increase engagement in the training activity; and finally, the link between appearance and emotional connection to the avatar.

1.5 Stylization

Stylization is the act of representing an object in a non-natural form [15]. Emoticons are an extreme example of stylization. The stylization of avatars, specifically virtual humans or virtual characters, can range from more realistic to cartoonlike in stylization [16]. Stylization might be done to accentuate specific features such as facial expressions, or it can enhance the ambiance of the situation (See Fig. 3). Japanese Anime is a good example of stylized characters, with larger-than-life, expressive eyes and almost childlike expressions [17]. The level of realism or stylization of a character is an important design decision that effects development time, cost, and user perceptions. Design decisions, such as this, need to be made early within the development process to ensure consistency with other features such as the environment [18]. In addition, the specific learning task must be considered to determine the desired level of realism of interpersonal interactions to support decisions of a learning population [19].

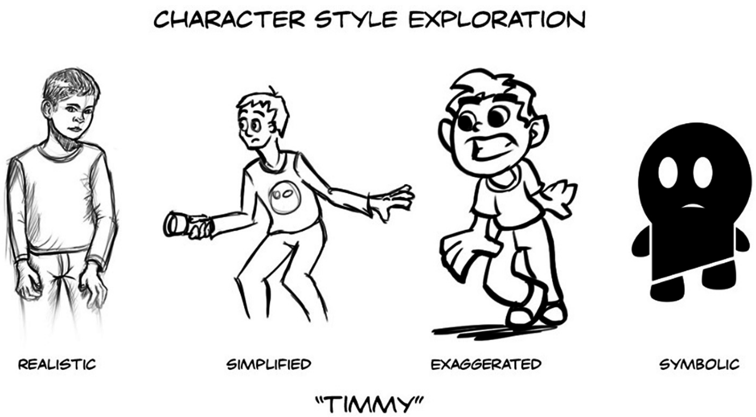


Fig. 3. Artist's rendition of character stylization [27]

2 Stakeholder Goals

2.1 U.S. Army Research Laboratory Simulation and Training Technology Center (ARL STTC)

The Army Research Laboratory's Simulation and Training Technology Center has the mission to execute the Army's science and technology program for simulation and training to accelerate learning and optimize human performance. The mission of this team within ARL STTC is to explore strategies to improve soldier training by leveraging commercial game technology. Puppeteering is a rapidly developing technology area that can provide functionality that AI is not yet able to provide within virtual environments.

2.2 Impact to the Army

Puppeering has the potential to support a wide range of training capabilities within the Army. Specifically, the U.S. Army Warfighters' Science and Technology Needs Bulletin, published by the U.S. Army Capabilities Integration Center at the U.S. Army Training and Doctrine Command (TRADOC) [20], that describes the need for virtual humans that should: (1) be capable of supporting "natural language processing to enable interactions (verbal and non-verbal);" (2) A virtual human must be able to "understand, reason and make assumptions about the environment;" and (3) "Virtual humans will populate large-scale simulations to expand the range of on-demand, interactive training opportunities and reduce human overhead support." (Pg. 40) While the goal is to have much of this functionality available via AI, that technology is currently not yet able to support these needs. Puppeteering is expected to bridge the gap until AI can more seamlessly meet the need.

The U.S. Army Learning Concept for Training and Education (TRADOC Pamphlet 525-8-2) [21] describes the future Army learning environment in this way:

> "Replicating the complex global environment within the learning context and conditions is critical to providing tough and realistic training and education. This complex global environment involves operations among human populations, decentralized and networked threat organizations, information warfare, and true asymmetries stemming from unpredictable and unexpected use of weapons, tactics, and motivations across all of the training domains. Adversaries are likely to employ information warfare to degrade mission command capabilities or conduct global perception management and influence campaigns. Army training and education must account for these and other factors during training and education activities. Adaptability is paramount; the learning system must provide training and education solutions to teams, Soldiers, and Army civilians synchronized to the operational tempo. To meet these challenges, Army training and education must do the following:
>
> (1) Portray the complex environment to develop leaders, Soldiers, Army civilians, and teams that understand the situation, apply appropriate judgment, adapt to changing conditions, and transition effectively between operations. Army training and education prepares Soldiers, and Army civilians to exercise mission command to exert influence on key individuals, organizations, and institutions through cooperative and persuasive means.
> (2) Create situations allowing individuals and teams to master fundamentals and hone skills.
> (3) Present complex dilemmas forcing leaders to think clearly about war to match tactical actions with operational and strategic objectives.
> (4) Create situations allowing individuals and teams to experience, become comfortable, and eventually thrive in ambiguity and chaos and then provide meaningful feedback on their performance.
> (5) Provide the required repetition, under the right conditions and with the right level of rigor, to build mastery of both fundamental and advanced warfighting skills (Pg. 11)."

The Army currently conducts live training events that, in some cases, require paid actors to interact with soldiers as part of a training scenario to establish a sense of realism. Soldiers are able to study the patterns of certain individuals to look for anomalies. Then they might engage various individuals in a mock-up of an operational environment. For example, they may meet with a key leader or law enforcement. These live training events can be costly and logistically complex. Similar training within a virtual environment could significantly reduce the overall cost of training for the Army. Characters could be controlled by AI until a soldier approaches one, then an actor/actress could step into that role and provide a rich interaction. A small number of role-players, maybe one female, one male, and possibly a native speaker or two, could hop into specific characters. Cost savings are important, but it cannot be at the expense of training capability. This represents just one use-case of how this technology can benefit Army training.

2.3 Defense Equal Opportunity Management Institute (DEOMI)

The Defense Equal Opportunity Management Institute (DEOMI) was established by the Department of Defense (DoD) in 1971 as the Government's premier institution for education, training, and research in human relations, equal opportunity, equal employment opportunity, and diversity. DEOMI is responsible for training Equal Opportunity Advisors (EOAs) for all branches of the uniformed services and civilian Equal Employment Opportunity (EEO) counselors for civil servants. Additionally, DEOMI provides policy and strategy guidance to the DoD for equal opportunity, sexual harassment, and sexual assault.

2.4 Impact to DEOMI

Building on previous research [22], this project will serve as the initial phase of a test plan to determine the usability of this platform as a DEOMI training tool. The scope of this investigation is to determine how realistic avatars are perceived by DEOMI students. The long-term goal of this project is to shape the future of EOA/EEO training by developing an immersive training environment for DEOMI facilitators and students. We aim to build a platform that overcomes the limitations of static multimedia training materials (e.g., training videos, computer programs), is capable of real-time adaptive flex-training, and provides comprehensive play-by-play verbal, postural, and interpersonal engagement feedback. Such a platform will expand DEOMI's capabilities, creating more robust and comprehensive programs.

3 Method

3.1 Participants

Participants will include 200 active duty Service Members enrolled in the Equal Opportunity Advisor Course (EOAC) or reserve Service Members enrolled in the Equal Opportunity Advisor Reserve Component Course (EOARCC) at DEOMI. Participants will receive credit that can be used to partially fulfill the EOAC/EOARCC requirement to complete volunteer hours.

3.2 Experimental Design

Using a within-subjects design, the present study will manipulate avatar stylization to determine how stylization influences perceptions of avatar realism. Participants will watch two videos, each with differently stylized avatars, welcoming them to the program they will be attending at DEOMI. After each video, participants will answer questions that will measure perceived avatar realism.

3.2.1 Avatar Videos

The team created two videos of a single avatar welcoming the EOAC/EOARCC students to DEOMI. To determine what avatar characteristics will result in the highest ratings of perceived realism, the stylization of the avatars were varied between the two videos such that one contained a stylized avatar (See Fig. 4) and the other contained an avatar with more realistic features (See Fig. 5).

Fig. 4. Stylized avatar male civilian

The script (Table 1), background, and all other aspects of the videos are identical. In the videos, the avatar is standing up in an office-like setting, making the entire body of the avatar visible to participants.

This will allow all aspects (e.g., facial expressions, posture, hand gestures) of the avatar to be rated for realism. In the video, the avatar is facing towards the viewer (perceivably making eye contact) and verbally delivers a message containing the welcoming statement, what they can expect to learn from the course, and general instructions on when/where to show up for class. Participants will watch both versions of the welcome message. The presentation will be counter-balanced to eliminate the possibility of confounding order effects.

Fig. 5. Stylized uniformed female

Table 1. Script stated to participants within study

"I want to personally welcome you to the Defense Equal Opportunity management Institute, the best equal opportunity training institute in the world. This course is rigorous and will require your full effort and attention. Frequently, personal experiences while at DEOMI are life changing. Upon completion, you will be poised to make positive contributions to your Service as an Equal Opportunity Advisor, Command Climate Specialist, or Program Manager.

We ask that you report to Patrick Air Force Base Lodging today NLT 1800 for check-in and in-processing. Please ensure you bring the following required items: copy of orders and DD Form 1610. My staff and I will work diligently to assist your check-in and resolve any issues that may arise.

Please ensure that you read the welcome packet in its entirety as each document has valuable information. Once again, welcome to DEOMI and congratulations on your selection as a member of this course. We look forward to assisting you in your continuing development as a world class Equal Opportunity professional."

3.2.2 Perceived Realism

At the conclusion of each video, participants will rate each avatar's realism. We will measure realism across two sub-dimensions: visual and behavioral realism (See Fig. 6).

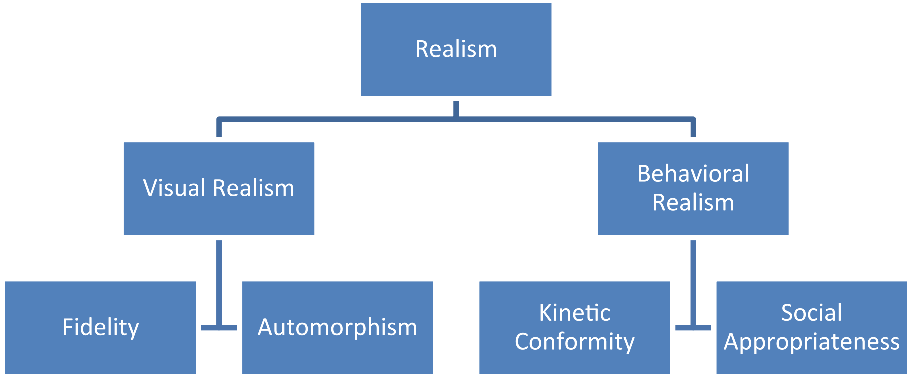


Fig. 6. Taxonomy of avatar characteristics

Within the sub-dimension of visual realism, we will measure fidelity and anthropomorphism. Fidelity refers to the quality of the avatar image and its surroundings. Anthropomorphism refers to the extent in which the avatar truly represents a human-like figure. Within the sub-dimension of behavioral realism we will measure kinetic conformity and social appropriateness. Kinetic conformity refers to the extent to which the avatar's movement resembles that of a real human being. Social appropriateness refers to the extent to which the verbal and nonverbal avatar responses align with communicative norms.

3.2.3 Analyses

Paired-sampled *t*-tests will be conducted to determine which of the two styles of avatars appear more realistic to the participants.

4 Way Ahead

Moving forward with this study we will be investigating how the characteristics and look of an avatar will effect participant's opinion and attitude. We will also examine what effect a virtual character controlled by a puppeteer produces in relation to how participants perceive emotions or an emotional connection within a training atmosphere.

Future research will build upon this work to explore the role bias plays in EEO tasks and interactions.

5 Conclusion

Technology and research into the topic of using a real-time human-controlled virtual character for EEO purposes is in its infancy. The research described in this paper is seeking to expand upon the listed study findings to see if stylization of virtual characters as compared to more photo-realistic characters produce a noticeable effect of how participants perceive emotions and how they build rapport with a puppeted avatar within a training atmosphere. This paper has described planned research and a projected plan where similar research might go to further.

References

1. Griffith, T., Dwyer, T., Kinard, C., Flynn, J.R., Kirazian, V.: Research on the use of puppeteering to improve realism in army simulations and training games. In: Lackey, S., Shumaker, R. (eds.) VAMR 2016. LNCS, vol. 9740, pp. 386–396. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39907-2_37
2. von der Putten, A.M., Krämer, N.C., Gratch, J., Kang, S.-H.: "It doesn't matter what you are!" explaining social effects of agents and avatars. Comput. Hum. Behav. **26**, 1641–1650 (2010)
3. McDonnell, R., Kokkinara, E.: Animation realism affects perceived character appeal of a self-virtual face. In: MIG SIGRAPH, Lisbon (2015)
4. Wallraven, C., Bülthoff, H.H., Cunningham, D.W., Fischer, J., Bartz, D.: Evaluation of real-world and computer-generated stylized facial expressions. Visual Computer (ICCAS), p. 24, November 2007
5. Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., Bailenson, J.N.: Immersive virtual environment technology as a methodological tool for social psychology. Psychol. Inq. **13**, 103–124 (2002)
6. Bhavani, D.K.: Capture artist graham qually: within the mocap zeitgeist. In: Motion and Performance Technology, 29 January 2018
7. Sharma, A., Agarwal, M., Sharma, A., Dhuria, P.: Motion capture process, techniques and applications. Int. J. Recent Innov. Trends Comput. Commun. **1**(4), 251–257 (2013)
8. Byron, K.: Carrying too heavy a load? The communication and miscommunication of emotion by email. Acad. Manag. Rev. **33**(2), 309–327 (2008)
9. Friedman, R.A., Currall, S.C.: Conflict escalation: dispute exacerbating elements of e-mail communication. Hum. Relat. **56**(11), 1325–1347 (2003)
10. Kruger, J., Epley, N., Parker, J., Ng, Z.W.: Egocentrism over e-mail: can we communicate as well as we think? J. Pers. Soc. Psychol. **89**(6), 925 (2005)
11. Schacter, D.L., Gilbert, D.T., Wegner, D.M.: Psychology, 2nd edn. Worth, New York (2014)
12. Eckman, P., Friesen, W.V., Scherer, K.: Body movement and voice pitch in deceptive interaction. Semiotica **16**, 23–27 (1976)
13. Penorroja, V., Orengo, V., Zornoza, A., Sanchez, J., Ripoll, P.: How team feedback and team trust influence information processing and learning in virtual teams: a moderated mediation model. Comput. Hum. Behav. **48**, 9–16 (2015)
14. Training Today: The Most Effective Training, 31 October 2017. http://trainingtoday.blr.com/article/most-effective-training-techniques/

15. Taubert, N., Loffler, M., Ludolph, N., Christensen, A., Endres, D., Giese, M.A.: A virtual reality setup for controllable, stylized real-time interactions between humans and avatars with sparse gaussian process dynamical models. In: ACM Symposium on Applied Perception (2013)
16. McDonnell, R., Breidt, M., Bulthoff, H.H.: Render me real?: investigating the effect of render style on the perception of animated virtual humans. ACM Trans. Graph. **31**, 91:1–91:11 (2012)
17. Levi, A.: The sweet smell of Japan: anime, manga, and Japan in North America. J. Asian Pac. Commun. **23**, 3–18 (2013)
18. Kapp, K.M.: Managing the gamification design process. In: The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education, pp. 212–215. Wiley (2012)
19. Qu, C., Brinkman, W.-P., Ling, Y., Wiggers, P., Heynderickx, I.: Human perception of a conversational virtual human: an empirical study on the effect of emotion and culture. Virtual Real. **17**, 1–15 (2013)
20. U.S. Army Training and Doctrine Command: The Warfighters' Science and Technology Needs Bulletin. U.S. Army (2017)
21. Training and Doctrine Command: The U.S. Army Learning Concept for Training and Education (2017)
22. Kang, S.H., Watt, J.H.: The impact of avatar realism and anonymity on effective communication via mobile devices. Comput. Hum. Behav. (2013)
23. Maxwell, D.B., Griffith, T.S., Finkelstein, N.M.: Use of virtual worlds in the military services as part of blended learning strategy. In: Handbook of Virtual Environments, pp. 959–1000. CRC Press-Taylor & Francis, Boca Raton (2014)
24. Griffith, T.S., Ablanedo, J., Nenneman, M.: Taking it to the EDGE. In: Proceedings of the International Defense and Homeland Security Simulation Workshop, Bergeggi, Italy (2015)
25. Hobby Consolas: Análisis de Assassin's Creed Origins the Hidden Ones, 25 January 2018. https://www.hobbyconsolas.com/reviews/analisis-assassins-creed-origins-hidden-ones-186668
26. Stern, M.: Give Andy Serkis an Oscar Nomination Already, 13 July 2017. https://www.thedailybeast.com/give-andy-serkis-an-oscar-nomination-already?ref=scroll
27. Brown, T.R.: Artist, Character Style Exploration [Art] (2013)
28. Epic Games: Unreal Engine, 23 October 2017. https://www.unrealengine.com/en-US/blog

Virtually Empathetic?: Examining the Effects of Virtual Reality Storytelling on Empathy

EunSeo Bang(✉) and Caglar Yildirim

State University of New York at Oswego, Oswego, NY 13126, USA
{ebang,caglar.yildirim}@oswego.edu

Abstract. Virtual reality is gaining attention as a new storytelling tool due to its ability to transport users into alternative realities. The current study investigated whether VR storytelling was a viable intervention for inducing a state of empathy. A short documentary about a prison inmate's solitary confinement experiences, *After Solitary*, was shown to two groups of participants. One group watched the documentary on a commercial VR headset (Oculus Rift) and the other group on a desktop computer via a YouTube 360° video. Results indicated the two groups did not differ in their state empathy levels and in their sense of presence levels. This suggests that watching the documentary in VR was not substantially different from watching it on YouTube with respect to the extent to which an individual empathizes with the emotional experience of another person.

Keywords: VR · Empathy · 360° video · Affective empathy · Cognitive empathy State empathy · Trait empathy · Immersion · Presence · Inmate rehabilitation

1 Introduction

Over the past few years, we have witnessed the rapid proliferation of commercially available virtual reality (VR) headsets. According to International Data Corporation [1], the market size of VR and augmented reality (AR) technology was 6.1 billion US dollars in 2016. The VR/AR market is also expected to increase to 143.3 billion US dollars by 2020. Similarly, augmented, virtual and mixed reality headset sales have dramatically increased worldwide and are estimated to increase for a few more years [2].

Accordingly, this new form of media has been gaining increasing attention as not just a tool of entertainment, but as a research instrument, too. VR has been utilized in various domains and its applications have been widely used. For instance, it has been used to provide psychotherapy for patients with eating disorders to help them overcome body image distortions [3], to reduce social anxiety symptoms [4], and to induce positive mood changes [5]. Unlike traditional media, when users are in virtual environments (VE), they often experience a sense of being in that environment, or presence [6]. Users often experience the feeling of entering into the virtual environment, and getting "physically situated in another." [7] This, often hard to distinguish from the concept of immersion, which is a measures "the extent to which the computer displays are capable of delivering an … illusion of reality to the sense of a human participant." [6]. Although

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 290–298, 2018.
https://doi.org/10.1007/978-3-319-91581-4_21

a clear definition and distinction between immersion and presences are yet to be well-defined, it is generally accepted that immersion are the necessary for experiencing presence [7], and two are closely intertwined with each other.

1.1 Empathy

One of the main reasons why VR is being increasingly used in research studies is that it affords the ability to transport users into an alternative reality that is different from their actual reality. Whether this transportation involves going to Mount Everest, watching the night stars in one's living room using Google Earth, or turning oneself into a three-legged alien, it allows users to go beyond physical boundaries of the present environment and to step into the realm of experiencing an alternative, virtual realities firsthand. Thus, the immersive nature of VR allows for having users put themselves in someone else's shoes and approach situations from their perspectives. This is one reason why VR experiences are a viable tool for inducing empathy – the ability to experience what others are experiencing. It transforms mere low-level sympathy- the ability to understand, as an abstraction, where another person is coming from into true empathy, experiencing second-hand what the other has felt.

Although there is not a single, commonly agreed-on definition of empathy, it can be defined as an individual's ability to understand and share another individual's feelings [8]. As the definition of empathy indicates, empathy is generally conceptualized as a multidimensional construct containing both sides of the coin: understanding another individual's situation or feelings, referred to as cognitive empathy, and feeling for another individual's emotions, referred to as affective empathy. Shen [9] argues that there is also associative empathy, which is concerned with the extent to which an individual identifies with how another individual feels. This third component is considered as a dimension of empathy by other researchers as well [4, 5]. It is argued that associative empathy serves a social function helping an individual establish social relationships with others [12]. Nonetheless, associative empathy has not been research extensively and is harder to define [9].

When empathy is being examined as a phenomenological construct, a distinction between state and trait empathy is usually highlighted. State empathy is when people experience some emotionally empathetic moments in response to another individuals' feelings. State empathy is situational and specific to its subject and involves "automatic and somatic responses" to another person's feelings [13]. State empathy can be distinguished from trait empathy, which refers to an individual's dispositional tendency to feel empathetic toward others' experiences in general [14]. For example, it is generally considered that women are more likely to demonstrate greater levels of trait empathy, when compared to men.

Due to the immersive nature of virtual environments and situational aspects of empathy, VR technology has been used to prime users to have a change of the heart on certain topics. Specifically, VR was used to help the caretakers of Alzheimer's patients have a better understanding of how it feels to experience dementia [15], to alleviate levels of racial bias by having Caucasian women see themselves in a darker colored skin [16], to induce empathy for those who had experienced war by putting the viewers in a disaster [17], and to raise awareness on the importance of the preservation of nature by having users grow from a little seed to a full grown tree, just to be cut down by humans [18].

Wijma et al. conducted an experiment in which participants who were primary informal caregivers of an Alzheimer's patient learned more about the disease in a VR environment [15]. In a pre/post-test design, participants watched a 360° simulation video in VR from a first-person perspective putting them in the shoes of a patient suffering from dementia. In one video, participants were immersed in a situation where it was the patient's birthday and everyone was celebrating and having a piece of birthday cake. The patient, however, had no clue why the surrounding people were celebrating and eating the cake. Participants provided self-reports of empathy among other measures. Results indicated positive changes in empathy ratings, demonstrating that the VR intervention helped induce a state of empathy. Participants also reported having a more in-depth understanding of what it is like to experience dementia.

Maister et al. investigated whether inducing ownership over a body avatar of a different race in VR affected implicit racial bias [16]. Participants, composed of Caucasian women, interacted with a body-tracking VR environment in which their virtual body was shown in different skin (dark-skinned, light-skinned, and purple). Results indicated that participants who had a dark-skinned virtual body demonstrated a substantial decrease in implicit racial bias [19], compared to the participants who had a purple virtual body. This study provides further evidence for the notion that VR can lead to state-like changes in an individual's perspective taking ability and thus induce a state of empathy.

The utility of VR as an intervention technology to induce a state of empathy can be empowered and bolstered by the extent to which the VR environment provides an immersive reality and leads to feelings of being psychologically present in the environment. In relation to VR, immersion refers to the objective quality of sensory input, pertaining to the extent to which visual graphics, sounds, haptic feedback, etc. feels real [7]. Presence, on the other hand, refers to the "subjective experience of being in one place or environment, even when one is physically situated in another" [7]. Although Schubert et al. argues that presence and immersion do not have a one-to-one relationship [20], it is commonly considered that sense of presence rises from immersion. Both immersion and presence are key factors in measuring the effectiveness of a VR environment.

Given the potential of VR to enable users to put themselves into others' shoes and the increasing use of VR as a storytelling tool, the current study investigated whether storytelling in VR could induce changes in an individual's empathy level toward a character in a VR environment. Participants watched a documentary, titled After Solitary by Frontline, about a former prison inmate's experience in solitary rooms on either a VR headset or a desktop computer. Following the documentary, participants provided self-reported ratings for their empathy level toward the inmate as well as for immersion and presence. The research question guiding the current study and the hypotheses were as follows:

Is VR storytelling a viable intervention for inducing a state of empathy?

- Hypothesis 1: Participants in the VR condition would report greater levels of immersion and presence as measured by self-reported ratings, when compared to those in the desktop condition.
- Hypothesis 2: Participants in the VR condition would report greater levels of state empathy as measured by self-reported ratings, when compared to those in the desktop condition.

2 Method

We used a between-subjects experimental design (VR vs. YouTube 360° video) to eliminate carry over effects from one condition to another. The dependent variables were self-reported sense of presence and state empathy ratings. What follows is a detailed description of the experimental method in terms of participants, materials, and procedure.

2.1 Participants

The sample consisted of 44 students (15 women, 29 men), with an average age of 22.35 ($SD = 3.49$). Participants received extra credit as compensation for their participation. The study was approved by the local human subjects committee.

2.2 Materials

Documentary. As pointed out earlier, the documentary used in this experiment was titled "After Solitary" by Frontline. The documentary is available as both VR documentary and YouTube 360° video [21]. The two versions are identical, except for the medium. Both versions lasted for 10 min. The documentary puts viewers in the shoes of a former prison inmate, Kenny Moore, who had been frequently placed in a solitary room. In the VR documentary, users can interact with the room the inmate was in and experience what prison inmates' life would be like. In the YouTube version of the same documentary, participants can still explore the room using the computer mouse. YouTube version also supports 360° viewing angle [21].

State Empathy Questionnaire. State empathy was operationalized as the average score on the State Empathy Questionnaire (SEQ), developed by Shen [9]. The SEQ consists of twelve items rated on a 7-point Likert scale. Four items tap into affective empathy, four into cognitive empathy, and four into associative empathy. The combined score in these three categories was used as a measure of state empathy. Higher SEQ scores indicated greater levels of state empathy.

Presence Questionnaire. Sense of presence was operationalized as the average score on the Presence Questionnaire, developed by Witmer and Singer [7]. The PQ consists of twenty four items rated on a a seven-point Likert scale to evaluate the level of presence experienced by the respondent. Four items were excluded from the original PQ. The adapted version consisted of twenty items The average score on all of the items was used as a measure of presence, with a higher score indicating higher degree of presence.

2.3 Procedure

Participants were randomly assigned into one of the two experimental conditions, VR documentary on Oculus Rift VR headset with no controllers or YouTube 360° video on a desktop computer. Before the experiment the participants completed online questionnaires which included an immersive tendency questionnaire (rated on a seven-point Likert scale), trait empathy questionnaire (rated on a five-point Likert scale), and some demographic questions regarding the individual's prior experience with VR, and political views (5-point scale ranging from very conservative to very liberal, with the option of no opinion). After the completion of the online screening questionnaire, participants were to make an appointment for when they could participate in the experiment. The screening questionnaire was used to ensure that the two groups did not significantly differ from one another in these variables.

Upon arrival in the lab, participants were provided informed consent and were encouraged to ask for clarifications. Then participants were assigned into their randomly selected groups. Both groups received the same set of instructions and in both cases participants put on a headphone to listen to the documentary while watching it. Participants watched the video in a private area and were instructed to inform the experimenter upon finishing the video. Once the video was finished, participants completed the post-experiment questionnaires on the same computer, which also took about 10 min. The post-experiment questionnaires consisted of the state empathy scale, some immersive tendency questionnaire items, and presence questionnaire.

3 Results

An independent samples *t* test was used to test two hypotheses regarding the between-group differences in presence ratings (hypothesis 1) and in state empathy levels (hypothesis 2). The alpha level was set to .05. Table 1 provides a summary of statistical analysis.

Table 1. Summary of statistical analysis

	n	*M (SE)*	*t*	*df*	*p*
Presence			.083	42	.93
VR headset	22	4.99 (.168)			
360° video	22	4.97 (.142)			
State empathy			−.014	42	.99
VR headset	22	5.04 (.162)			
360° video	22	5.05 (.225)			

In relation to hypothesis 1, results showed that there was no significant main effect of method of watching the documentary on presence ratings, $t(42) = .083$, $p > .05$. Participants who watched the documentary via the YouTube 360° video ($M = 4.97$, $SE = .142$) and participants who watched the documentary on Oculus Rift ($M = 4.99$, $SE = .168$) did not significantly differ from one another in their self-reported presence levels. See Fig. 1.

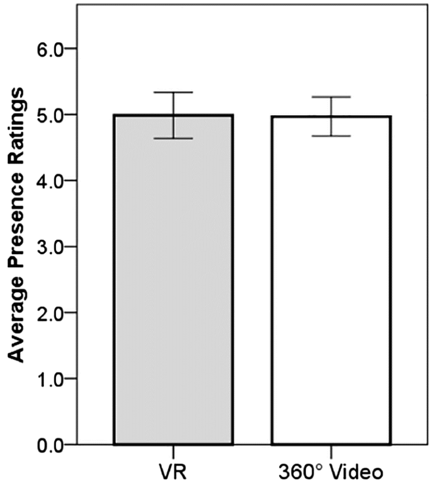


Fig. 1. Bar graph showing average presence ratings as a function of condition.

In relation to hypothesis 2, results showed that there was no significant main effect of method of watching the documentary on state empathy levels, $t(42) = -.014$, $p > .05$. Participants who watched the documentary via the YouTube 360° video ($M = 5.04$, $SE = .162$) and participants who watched the documentary on Oculus Rift ($M = 5.05$, $SE = .225$) did not significantly differ from one another in their self-reported state empathy levels. See Fig. 2.

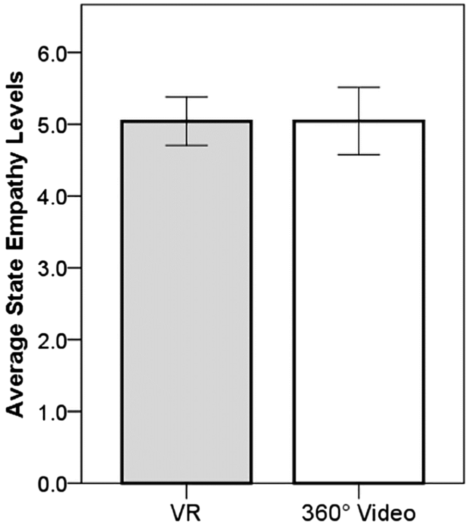


Fig. 2. Bar graph showing average state empathy scores as a function of condition.

4 Discussion

The current study investigated the viability of VR storytelling as an intervention to induce changes in state empathy levels. The results provided no support for the hypothesis that participants in the VR condition would report greater levels of immersion and presence as measured by self-reported ratings, when compared to those in the desktop condition. There was no supporting evidence for the second hypothesis that participants in the VR condition would report greater levels of state empathy as measured by self-reported ratings, when compared to those in the desktop condition. Taken together, these results indicate that the two groups did not differ in their presence and state empathy levels, contrary to prediction. This finding suggests that watching the documentary in VR was not substantially different from watching it on YouTube with respect to the extent to which an individual empathizes with the emotional experience of another person.

The results from the current experiment are contradictory to prior studies providing corroborating evidence for the effectiveness of VR in inducing changes in state empathy [15–17]. That said, the conceptualized mediating effect of sense of presence in a VR environment on state empathy could explain the null findings from the current experiments. Hypothesis 1 was concerned with the differences in sense of presence and was not supported by the data. Thus, the fact that participants did not differ in their sense of presence may explain why they did not differ in their state empathy levels either. This interpretation is in line with prior work on the diegetic effect [21]. Schubert et al. pointed out that sense of presence is not commonly reported in traditional media, except for diegetic effect. It is possible that participants may experience certain emotional changes while interacting with a VR environment, despite being unable to consciously report it

[22]. Hence, it could be argued that participants in the VR condition experienced greater changes in their state empathy at the physiological levels but their perceived sense of presence was similar to that of the YouTube 360° video condition. Since the current study employed no physiological measures, this potential explanation remains as an open question for future research.

One limitation of the current study is related to the content of the documentary that was used in the experiment. As pointed out before, the documentary was about a former prison inmate's experience with solitary confinement. The sensitive nature of the topic and its content may be the reason why the two groups did not differ from one another in state empathy levels. Future studies could address this limitation by employing other documentaries that are less sensitive in nature. Additionally, owing to its complex operational definition, measuring state empathy in VR environments may not be a clear-cut process. Thus, state empathy might be difficulty measure using self-reported measures only, without physiological measures. Future research could attempt to develop physiological measures of state empathy. Also, the current experimental design could be replicated to investigate whether VR storytelling is a viable intervention for inducing changes in other emotions that have an establish valid physiological and behavioral measures (e.g., measurement of enjoyment of a comedy show by measuring how much participants laugh or by electrodermal activity).

Acknowledgments. The study was conducted in the SUNY Oswego VR First Lab, established through the academic partnership between VR First and SUNY Oswego. The authors would like to thank VR First for supporting the establishment and maintenance of the lab.

References

1. IDC: Worldwide Revenues for Augmented and Virtual Reality Forecast to Reach $162 Billion in 2020, According to IDC. http://www.idc.com/getdoc.jsp?containerId=prUS41676216
2. Merrill Lynch: Theme Watch (YA 2017: Year of Tech Disruption, Earth, & Millennials Themes). https://olui2.fs.ml.com/publish/content/application/pdf/GWMOL/Theme-Watch-2017-Year-Ahead.pdf
3. Gutiérrez-Maldonado, J., Ferrer-García, M., Caqueo-Urízar, A., Moreno, E.: Body image in eating disorders: the influence of exposure to virtual-reality environments. Cyberpsychol. Behav. Soc. Netw. **13**, 521–531 (2010)
4. Kampmann, I.L., Emmelkamp, P.M.G., Hartanto, D., Brinkman, W.-P., Zijlstra, B.J.H., Morina, N.: Exposure to virtual social interactions in the treatment of social anxiety disorder: a randomized controlled trial. Behav. Res. Ther. **77**, 147–156 (2016)
5. Navarro-Haro, M.V., López-del-Hoyo, Y., Campos, D., Linehan, M.M., Hoffman, H.G., García-Palacios, A., Modrego-Alarcón, M., Borao, L., García-Campayo, J.: Meditation experts try Virtual Reality Mindfulness: a pilot study evaluation of the feasibility and acceptability of Virtual Reality to facilitate mindfulness practice in people attending a Mindfulness conference. PloS One **12**, e0187777 (2017)
6. Slater, M., Wilbur, S.: A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. Presence Teleoperators Virtual Environ. **6**, 603–616 (1997)

7. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence **7**, 225–240 (1998)
8. Hogan, R.: Development of an empathy scale. J. Consult. Clin. Psychol. **33**, 307–316 (1969)
9. Shen, L.: On a scale of state empathy during message processing. West. J. Commun. **74**, 504–524 (2010)
10. Decety, J., Jackson, P.L.: The functional architecture of human empathy. Behav. Cogn. Neurosci. Rev. **3**, 71–100 (2004)
11. Decety, J., Jackson, P.L.: A social-neuroscience perspective on empathy. Curr. Dir. Psychol. Sci. **15**, 54–58 (2006)
12. Campbell, R.G., Babrow, A.S.: The role of empathy in responses to persuasive risk communication: overcoming resistance to HIV prevention messages. Health Commun. **16**, 159–182 (2004)
13. Preston, S.D., de Waal, F.B.M.: Empathy: its ultimate and proximate bases. Behav. Brain Sci. **25**, 1–20 (2001)
14. Davis, M.H.: Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. **44**, 113–126 (1983)
15. Wijma, E.M., Veerbeek, M.A., Prins, M., Pot, A.M., Willemse, B.M.: A virtual reality intervention to improve the understanding and empathy for people with dementia in informal caregivers: results of a pilot study. Aging Ment. Health. 1–9 (2017)
16. Maister, L., Slater, M., Sanchez-Vives, M.V., Tsakiris, M.: Changing bodies changes minds: owning another body affects social cognition. Trends Cogn. Sci. **19**, 6–12 (2015)
17. A War Survivor's Virtual Reality Film Brings the Terror of a Conflict Zone to Life – Creators. https://creators.vice.com/en_au/article/8qv7jg/giant-vr-film-gdc
18. Can virtual reality emerge as a tool for conservation?—Environment—The Guardian. https://www.theguardian.com/environment/2016/jun/28/can-virtual-reality-emerge-as-a-tool-for-conservation
19. Peck, T.C., Seinfeld, S., Aglioti, S.M., Slater, M.: Putting yourself in the skin of a black avatar reduces implicit racial bias. Conscious. Cogn. **22**, 779–787 (2013)
20. Schubert, T., Friedmann, F., Regenbrecht, H.: The experience of presence: factor analytic insights. Presence Teleoperators Virtual Environ. **10**, 266–281 (2001)
21. Official, FRONTLINE PBS: After Solitary 360°—FRONTLINE. https://www.pbs.org/wgbh/frontline/article/after-solitary/

The Role of Psychophysiological Measures as Implicit Communication Within Mixed-Initiative Teams

Kim Drnec[1(✉)], Greg Gremillion[2], Daniel Donavanik[2], Jonroy D. Canady[1], Corey Atwater[3], Evan Carter[1], Ben A. Haynes[4], Amar R. Marathe[1], and Jason S. Metcalfe[1]

[1] Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD, USA
kim.a.drnec2.ctr@mail.mil,
{jonroy.d.canady.civ,evan.c.carter3.civ,
amar.r.marathe.civ,jason.s.metcalfe2.civ}@mail.mil
[2] Adelphi Laboratory Center, Adelphi, MD, USA
gregory.m.gremillion.civ@mail.mil
[3] DCSCorp, Warren, MI, USA
catwater@dcscorp.com
[4] U.S. Army Tank Automotive Research Development and Engineering Center, Warren, MI, USA
ben.a.haynes2.civ@mail.mil

Abstract. There has been considerable effort, particularly in the military, at integrating automated agents into human teams. Currently, automated agents lack the ability to intelligently adapt to a dynamic operational environment, which results in them acting as tools rather than teammates. Rapidly advancing technology is enabling the development of autonomous agents that are able to actively make team-oriented decisions meaning truly intelligent autonomous agents are on the horizon. This makes the understanding of what is important to team performance a critical goal. In human teams, mission success depends on the development of a shared mental models and situation awareness. Development of these constructs requires good intra-team communication. However, establishing effective intra-team communication in a mixed-initiative team represents a current bottleneck in achieving successful teams. There has been significant research aimed at identifying modes of communication that can be used both by human and agent teammates, but often neglects a source of communication or information for the agent teammate that has been adopted by the human robot community to increase robot acceptance. Specifically, the use of psychophysiological features supplied to the agent that can then use algorithms to infer the cognitive state of the human teammate. The utility of using psychophysiological features for communication within teams has not been widely explored yet representing a knowledge gap in developing mixed-initiative teams. We designed an experimental paradigm that created an integrated human-automation team where psychophysiological data was collected and analyzed in real-time to address this knowledge gap. We briefly present a general background to human automation teaming before presenting our research and preliminary analysis.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 299–313, 2018.
https://doi.org/10.1007/978-3-319-91581-4_22

Keywords: Automation · Autonomy · Robot · Mixed-initiative teams
Human automation interaction · Bi-directional communication
Psychophysiology

1 Introduction

Automated agents, regardless of their instantiation, i.e., a computer or a robot, have become increasingly adopted by both the military and society in general. In the case of the military, automated agents are generally employed to carry out tasks that are too difficult or too dangerous for a Soldier [1]. For example, the MARCbot is used by the military to investigate suspicious objects, and can even look behind doors because of the placement of its camera [1]. Alternatively, automated agents in society are usually used for tasks that are repetitive or that require time-sensitive and complex computations. In both domains, however, the agent, currently, is largely used as a tool [1] to extend the human operator's functional 'reach', or to improve performance of a task, or both, as opposed to a teammate. Moreover, despite the near ubiquitous presence of automated agents, joint system (i.e., the human operator and the agent) performance has failed to reach desired expectations [2]. While many reasons for this failure have been forwarded [3–5], one of the most important identified by the Department of Defense [9] is that the human operator normally assumes a supervisory role. This role often places the human operator 'out of the loop' [6–8] resulting in reduced operator situation awareness of agent function and degree of task completion [9]. This supervisory role, which might include tasks such as monitoring a display has been necessary due to the inherent limitations of automated agents; they have not been capable of autonomous behavior, but rather they have often replaced a human activity in a brittle manner.

Recently, two key developments have occurred, largely in the field of robotics, that are likely to change the way that humans and automated agents interact to achieve a shared goal in fully collaborative situations. First, advances in artificial intelligence (AI) are rapidly providing automated agents with the ability to flexibly adapt behavior based on operator needs, or changing task demands. Although these abilities such as learning about a human partner or an environment are yet still relatively primitive, this technical achievement in AI enables truly autonomous agents, capable of making team-oriented decisions [10]. Secondly, there is emphasis being placed on fully integrating autonomous agents and humans, particularly in the military domain [9]. This full integration of agent and human results in mixed-initiative teams that will require genuine human-agent collaboration. These developments will essentially move automated agents from tools to teammates [10–12]. The purpose of this paper is twofold. First we broadly discuss the advances in theory and technology enabling the formation of successful mixed-initiative teams, focusing on intra-team communication. We then highlight a critical gap that is, as yet to be fully-addressed. That is, we focus both our literature review and research on how to improve intra-team communication through the use of implicit human state information derived from psychophysiological signals. Second, we describe our research that is aimed at filling that gap, and finally, we present very preliminary results from several individuals. In the next section, we enter our discussion by carefully delineating a line of thought that arrives at the concept of better human-agent teaming through shared mental model development as augmented by the application of physiology-based monitoring and prediction of human decisions during a joint human-automation driving task.

2 Moving Agents from Tool to Teammate Within Mixed-Initiative Teams

Within the human agent teaming (HAT) literature, the terms robot, autonomy, and automation are seldom defined and are occasionally conflated. For example, the Defense Science Institute in 2016 [9] held a workshop on bi-directional communication for human robot teaming, but within the report, autonomy is used both as an adjective to describe robot capabilities, and as a synonym for robots [9]. To avoid confusion in the following discussion we consider the space of (HAT) as being represented by the Venn diagram in Fig. 1. Further, we define the term automation to be any machine or software that performs an automated function that generally lacks the capability of independent action. We define an autonomy as an intelligent agent that may be capable of independent action, i.e., performing actions not previously hard coded. Therefore, a robot might be either an automation or an autonomy. However, as technology improves and increasingly sophisticated autonomies appear on the horizon, there will inevitably be some grey area between automation and autonomy. We use the term ‘agent’ to designate a robot, an autonomy, or an automation. Finally, we define a team as ‘a group of two or more people who interact dynamically, interdependently, and adaptively towards a common value or goal’ [13]. We use the term mixed-initiative teams as used in the literature [14], to designate a team comprising at least one human and one agent.

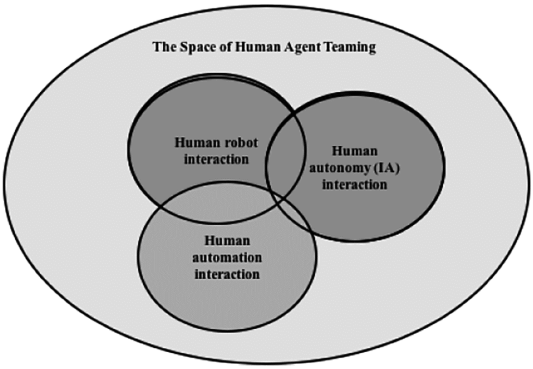


Fig. 1. The conceptual space of human agent teaming as discussed herein

Literature involving human only teams posits that successful mission completion heavily relies on the development of two main constructs. Teams need to have a shared mental model of the problem space, and shared situation awareness of evolving environments [10, 13, 15]. Mental models held by team members reflect an understanding of teammate intent during problem solving such that each team member can infer the causes of a teammate's behavior [16], and act symbiotically to achieve a goal. Shared situation awareness includes a common representation of the dynamically changing environment or problem. In human-human teams, these constructs are continuously being updated [12], and therefore it stands to reason that the ability for an agent teammate to update its mental model and situation awareness will be critical to successful mixed-initiative teams. Research efforts entailing mixed-initiative teams must focus on engendering these constructs effectively within the team. Therefore, team dynamics as a whole need to be directly addressed as opposed to the traditional model, where focus is usually directed toward improving human *or* agent performance [17], essentially treating teammates as independent actors rather than interdependent entities.

Effective communication between teammates, whether explicit or implicit, is the cornerstone of establishing accurate mental models and shared situation awareness within a team [10, 15, 17, 18]. Within this team construct, it is implied that the communication be bi-directional between the team members in order to be effective. For mixed-initiative teams, the requirement for bi-directional communication is even more important if an agent is to move from being a tool to a teammate [19]. However, this ability to have bi-directional communication, a key for team success, represents a bottleneck in furthering mixed-initiative team performance [15]. There are, therefore, considerable research efforts aimed at identifying effective methods of bi-directional communication for use in mixed initiative teams [11, 19–22]. This research must explore modes of communication that consider the nature and characteristics of each teammate, as well as any constraints imposed by the current operational environments.

Multimodal communication (MMC) has been identified as one method to facilitate successful bi-directional communication [18, 20]. MMC is communication using more than one modality (voice, gestures, etc.) [20]. The reported flexibility of MMC means that communication can be accomplished intuitively and efficiently to best suit the individual situation [20]. For example, gesturing allows for silent communication, and haptic communication [20] allows team members to communicate even when they are out of the line of sight of each other. MMC, however, while flexible, and intuitive to human teammates, explicitly centers on the five human senses [20]. This five-sense-centric model of bi-directional team communication ignores signals that could be readily interpreted by most agents, namely, psychophysiological signals, which can be acquired and processed online using contemporary signal processing tools [23].

Psychophysiological signals have been successfully used to infer affective state during human robot interactions. For example, it is known that rapid movements by robots tend to induce a degree of anxiety in humans [24, 25]. However, if a robot teammate identifies that the human is in a state of anxiety, and adjusts its movements appropriately, anxiety is reduced [24]. Similarly, inferences based on psychophysiological signals can allow the robot to dynamically alter its behavior. For instance, it can change its level of autonomy, reallocate tasks, and infer when, and how, to query the human teammate [21, 26]. A predominant of work in this domain, with notable exceptions [12, 15, 21], appears to be largely focused on two goals; either a robot *replacing* a human activity or towards making the robot more acceptable to the user, which engenders trust, long considered necessary for successful human agent teaming [27, 28]. However, little research has been directed at explicitly exploring the utility of incorporating implicit psychophysiological signals as a form of communication into mixed-initiative teams. Our research aims to address this gap.

For our experiment, we developed a research paradigm that has some of what we believe to be critical elements of a mixed-initiative team, even if not at the full scale of emerging mixed initiative teams that are of ultimate interest. The elements included in our research paradigm consisted of a participant, a simulated vehicle, an optional driving automation and a controller. All these elements were fully integrated into a control theoretic framework that approximated a team. We provided psychophysiological communication to the controller in real time with the intent that the controller use this implicit communication as data to predict driver behavior. If participant driver behavior was predicted to be sub-optimal the control system could suggest, through a visual display, what behavior might be more beneficial for successful task completion. We had three expectations that would approximate metrics of the utility of including psychophysiological communication to the agent. First, that in conditions with integrated psychophysiological communication, joint system task performance would be better than without this implicit form of communication. Second, that participant trust would be highest in conditions including psychophysiological communication to the controller. Finally, subjective participant workload measures would be reduced in conditions with the psychophysiological communication as compared to conditions without that communication.

3 Materials and Methods

3.1 Overview

We based this research on a leader-follower simulated driving paradigm, well explored by our lab [29], involving a semi-automated driving automation capable of lane and speed maintenance, but not collision avoidance. We contend that this is a natural context for observation of human behaviors in the presence of an automated agent, particularly as driving is a common task familiar to most adults. In addition to providing the participant with a static automation, i.e., non-adaptable automation *vis a vis* capabilities, we added a control system designed to help the participant make appropriate decisions about what agent (human or automation) should be in control of the simulated vehicle at a point in time given relative agent capabilities. Individual participant capabilities were sampled by the control system at the outset of the experiment, essentially establishing a type of ‘ground truth’ for participant ability, while the control system had *a priori* knowledge of the automation’s performance. In our paradigm, the control system functioned as an advanced automation, capable of assessing participant ability and identifying relevant environmental features such as a tight turn. The control system could then integrate that information with psychophysiological signals to produce team-oriented probabilistic decisions about what agent should be in control of the simulated vehicle. Further, the control system, was designed such that it could ‘act’ on the outcome of the calculation. It accomplished this by adjusting a visual display (described later), from herein termed the actuator, where the term actuator is defined as the mechanism by which the controller acted on its environment. The actuator indicated to the participant what agent the system believed was likely to perform better given the previously calculated 'ground truth', and the current environment. The controller utilized one of three different algorithms, depending on experimental condition, to control the behavior of the actuator. One of these conditions integrated psychophysiological features.

3.2 Facilities and Equipment

The equipment and methods are nearly identical to that of previous research and a detailed description of the paradigm is given by Metcalfe et al. (2017). The current research, described below, was conducted in a sound proofed experimental chamber at the Cognitive Assessment Simulation and Engineering Laboratory, Aberdeen Proving Ground (APG). The simulation was realized by three large monitors on a desktop that displayed the simulated environment provided by SimCreator (Realtime Technologies, Inc; Royal Oak, MI). Participants sat in a chair that resembled an actual driving seat. During the experimental tasks participants were outfitted with a suite of sensors for recording psychophysiological activity. Specifically, participants donned a Biosemi ActiveTwo (BioSemi BV; Amsterdam, Holland) system to enable recording electroencephalographic brain activity (EEG), electrocardiographic activity (ECG), electrooculargraph activity reflecting eye motion (EOG), and skin electrical conductance (electrodermal activity; EDA).

Data collected, processed and analyzed in real-time included psychophysiological (EEG, EDA, and vertical and horizontal eye movements (vEOG, hEOG)), behavioral (e.g. automation use, brake and throttle inputs), and variables related to the participant's vehicle (e.g. heading error, speed etc.). In addition to the psychophysiological data collected, participants completed a set of surveys including the Big Five Inventory for assessing personality traits [31], a demographic questionnaire, and after each condition surveys appropriate to the condition. After all conditions participants completed a NASA-TLX [32] to assess the level of subjective workload, and a simulator sickness questionnaire to assess motion sickness. Further, in conditions where the automation was available participants completed a system trustworthiness and display trustworthiness surveys. Behavioral and participant vehicle data were collected at 60 Hz and recorded by SimCreator.

3.3 Experimental Task and Design

In this leader follower task participants drove one and one-half laps around a two-lane simulated course with ambient traffic (Fig. 2 bottom). Task objectives included maintaining lane position, a "safe" distance from the lead vehicle, and avoiding collisions with ambient traffic and frequently appearing pedestrians. The automation could control speed and lane position, but had no collision avoidance capabilities. Pedestrians appeared approximately every 7 s, distributed randomly on either side of the road, and 15% stepped into the vehicle path. Participants were asked to respond to pedestrians using buttons on a game controller, which removed the pedestrian from the road, thereby avoiding potential collisions. For conditions in which the driving automation was available, participants had the option to enable or disable it at any moment. It could be disabled through application of the brake, throttle inputs, or an accelerator-pedal- adjacent toggle foot switch, and enabled by depressing the same toggle foot switch. Lateral and longitudinal perturbations were introduced to add additional challenge to the driving task. The lateral perturbations simulated gusts of wind that tended to push the participant vehicle out of the lane increasing the risk of collisions with ambient traffic and pedestrians. Longitudinal perturbations were implemented by increasing or decreasing the speed of the lead vehicle. Participants were provided with a visual display screen (Fig. 2 top) denoting current score (described later), consequence zones (described later), control agent (human or automation) indicated with a green chevron (Fig. 2 top image), and in three of the five conditions, a functioning actuator display Fig. 2, top image, black bar at the top of screen; described in detail later).

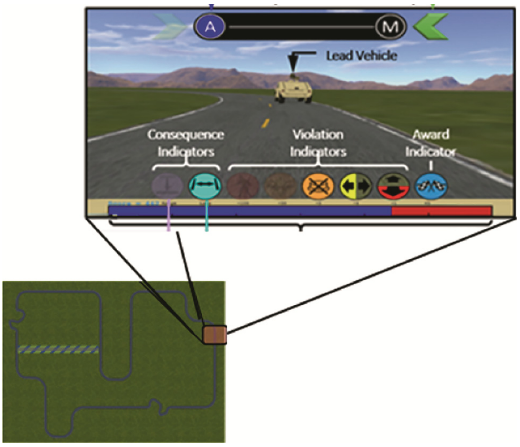


Fig. 2. At the top is a depiction of the visual display provided for the participants. At the top of the display is an indicator of what agent is in control. Here, the chevron points to ‘M’ (manual) so the participant is in control. At the bottom of the display is a sliding bar depicting point losses. The bottom Fig. illustrates the course the participants were to navigate.

The average drive time around the course (Fig. 2 bottom image) for each condition lasted approximately 18 min and consisted of 19 zones defined by environmental feature. For example, a section of course with an s-curve would be one zone, whereas another section of the course with a straight-away would constitute another zone. All zones were approximately equal in length. The purpose of the zones is to create circumstances wherein one agent would be preferred over the other in terms of performance. For instance, humans are generally superior at handling s-curves, whereas the driving automation was superior at handling straight-aways. In different conditions, zones would be assigned changing consequence designations for either lane or range violations and the task with an increased consequence was indicated by the visual display. For example, a zone might be designated high consequence for lane deviation meaning that, in that zone, penalties for lane deviations increased. There were four different sets of zone consequences; one for the first and last conditions, and three counterbalanced sequences for the other three conditions. For each experimental condition participants started with 500 points and lost points for each performance decrement. The point loss was proportional to real-life consequences, meaning that the most points were deducted for collision with a pedestrian, and the fewest points were deducted for lane and range violations. For zones denoted by increased consequence, point loss increased for that zone for the indicated performance decrement. Points remaining at the end of each condition were translated into monetary compensation with a simple algorithm. Point losses and current score were indicated by a sliding color bar (Fig. 2) at the bottom of the display screen.

Fig. 3. Informational icons provided in the display to the participant. At the top (A) are icons that represent possible performance decrements. These are greyed out (B) when there is no error, and are lit when the error occurs. In Fig. 3 B, illuminated icons indicate that there has been a collision with a pedestrian and that the wrong button has been pressed to clear the pedestrian from the road. At the far right are the consequence indicators. Here, in the bottom row (B) the lane deviation icon is illuminated signifying that for the current zone there is an increase in point losses for this error. Only one row of icons is displayed during the experiment; Fig. 3 A where all icons are illuminated is for illustrative purposes only.

Each experimental session consisted of five unique conditions; three conditions determined by the algorithm the controller used to control the behavior of the actuator, one manual condition, and one condition with no actuator present, but with the automation available. The first condition was always a 'manual' condition where the driving automation was not available, and the actuator did not function, although the actuator display indicated with a green chevron that the participant was in control of the vehicle. This condition served two important purposes. First, it is important for a participant to be able to gauge their ability to perform the task without aid from an automation. This is because experience indicates that whether a person uses an automation is highly dependent on their perceived ability to succeed at a task themselves versus their perception of an automation's ability to perform the same task [33, 34]. Second, the control system, to be able to generate relative probabilities of success for each agent for each zone, needed to have an accurate assessment of the participant's innate ability to perform the task relative to the known ability of the automation.

The other four conditions were as follows; no actuator (NA), unscaled actuator (UA), scaled actuator (CA), and psychophysiologically scaled actuator (PA). These actuator displays are shown in Fig. 4, and the method of counterbalancing conditions and consequence sequences is described below. The NA condition was a condition where the automation was available, but no indication was provided to the participant regarding probable success of either agent. The UA actuator algorithm determined the objective probability of each agent to succeed at the task in a zone. The CA actuator algorithm considered not only each agent's abilities, but the consequence designation of the zone. For example, if a zone with a tight s-curve was the current zone and it was designated as an increased consequence for range deviation the control system needed to consider that fact. The automation is generally better at range maintenance, and the human driver is generally superior at maneuvering tight curves. Therefore, if an increase consequence for range deviation in a tight curve was indicated the system might still suggest that the automation be in control instead of the participant. When active, the actuator indicated the preferred agent by presenting a colored ball that moved smoothly from 'A' to 'M' (Fig. 4) as the outputs of the algorithm were derived. The ball changed colors depending on the recommended agent. If the automation was recommended the ball turned blue as it moved towards the 'A'. Conversely, as the ball moved toward the 'M', suggesting that manual control was preferred, it turned red. It should be noted, however, that the actuator display in the UA, CA, and PA conditions appeared identical to the participant.

Fig. 4. Examples of the actuator display for CA, UA and PA conditions which were visually identical, differing only in the algorithm used to control their behavior. In the top left the automation is in control and the output of the control algorithm is leaning towards automatic control. In the top, right the automation is also in control, but the controller is leaning towards manual control. In the bottom left the automation is in control, and the controller is suggesting that this state be continued. In the bottom, right, the participant is in control and the controller is suggesting that they remain in control. (Color figure online)

The PA condition was introduced as an explicit test of the utility of providing implicit communication to an agent in the form of psychophysiological signals from the participant. The control algorithm that drove the actuator behavior in the PA condition considered the objective probability of the success for both agents, as in the UA condition. However, in the PA condition, the control algorithm also factored in inferences made from psychophysiological features about the likelihood of the participant activating or deactivating the automation at any given time. This ability was made feasible by previous work by our lab which demonstrated that a machine classifier, taking specific environmental and psychophysiological features as inputs, could predict when a participant was about to make a change in control authority with almost 70% accuracy. This classifier was implemented in real time in the control system, and the output was included in the algorithm driving the behavior of the actuator. In the PA condition the probability of one agent being in control versus the other was added to the objective probability of success for each agent in a principled way based on probability theory.

The purpose of the experiment was twofold, and one of the aims partially dictated experimental design. The primary purpose was to assess the utility of adding implicit communication through psychophysiological signals to the controller as measured by overall performance of the task. The second purpose was to act as a continuation of a previous experiment and essentially increase overall subject numbers. The earlier experiment which included the manual, CA, UA and NA conditions was conducted at the U.S. Army's Tank and Automotive Research Development and Engineering Center (TARDEC) in Michigan. Therefore, the CA, UA, and NA conditions that occurred in both experiments were always counterbalanced along with their three possible consequence sequences using a Latin Square, creating a 3×3 factorial design. For execution of the experiment described in this paper, the manual condition was always first for reasons described earlier. To allow for inferences regarding the effects of integrating psychophysiological signals into the control system, half of the participants (n=4) experienced the PA condition immediately after the manual condition and half (n=5) experienced it as the last of the five conditions.

4 Preliminary Results

Preliminary results are based on data available from an initial sample of 5 participants, although we have plans to collect data from a total of 18 participants. We originally considered that performance and trust (trust of the actuator display) would be highest in the PA condition. The inclusion of performance based metrics in our analysis has obvious motivation. If the metrics reflected high levels of performance, it might be inferred that the team in that condition was successful as compared to other conditions. We used points lost per second as a metric of performance (Fig. 5A) as a way of accounting for varying lengths of time on the course. Trust was considered as a relevant variable because of its presumed importance as to how an operator decides to use an automation. It is widely considered that if there is an appropriate level of trust (not too high and not too low) the automation will also be used appropriately. Figure 5B shows the mean trustworthiness of the display scores in the three conditions where an actuator display was available. We also expected that workload (Fig. 5C) would be reduced in the PA condition as compared to all five conditions. There was no statistical difference in performance, trust, or workload across the conditions, but there are tendencies seen on visual inspection of Fig. 5 that if they remain when a larger data set is analyzed, would make interesting results. For example, it appears that performance was improved overall by having the automation available (Fig. 5A) as compared with the manual condition. Further, trust (Fig. 5B) appears to be highest in the PA condition with a very narrow distribution, indicating that most participants felt they could trust the actuator display. Interestingly, workload also appears to be lowest in conditions where an actuator was available (UA, CA, PA), which might suggest that participants found the task easier when they were given cues as to which agent should be in control.

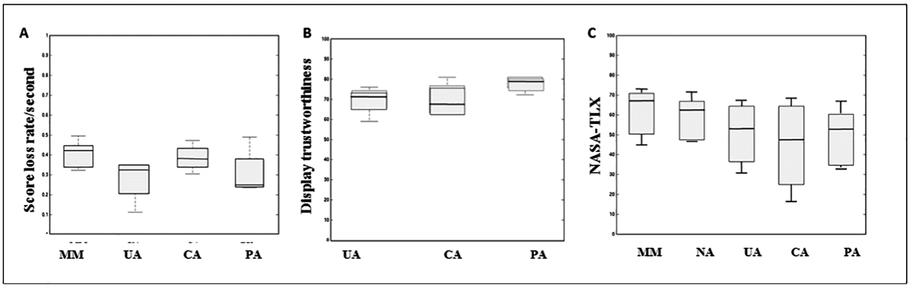


Fig. 5. (A) Score loss per second, (B) actuator display trustworthiness, (C) workload (unscaled NASA-TLX). MM (manual condition), UA (unscaled actuator), CA (scaled actuator), NA (no actuator), PA (psychophysiological actuator). The black line in the bars represents the mean value.

5 Discussion

We designed this experiment to determine if integrating implicit communication in the form of psychophysiological data would increase overall team performance, presumably by facilitating development of shared mental models and situation awareness. As noted, this is an important question that could significantly affect the design and future success of mixed-initiative teams. We formulated three general expectations that if confirmed, would support the utility of this form of intra-team communication. The first was that in conditions with integrated psychophysiological communication, joint system task performance would be better than without this implicit form of communication. The second expectation was that participant trust would be highest in conditions including psychophysiological communication to the controller. Finally, the third was that subjective participant workload measures would be reduced in conditions with the psychophysiological communication as compared to conditions without that communication.

Although there were no statistically significant results to report, we nevertheless can speculate about interesting observations regarding the behavior of performance, trust and workload across different conditions. The results seem to indicate that the use of physiological information within the controller has the potential to produce performance that is at least on par with other actuated conditions; according to our preliminary observations, display trustworthiness may tend to be improved and workload reduced when physiology-based state prediction was incorporated. If these results hold up with the addition of more data, it could be argued that there is a compelling reason to include psychophysiological signals. A reduction in subjective workload can affect how a human operator processes environmental stimuli that directly relate to the task [35]. If a human operator is experiencing a high level of workload, the operator is likely to not only have fewer cognitive resources to process important task-relevant information, but also she is likely to have a reduced situation awareness [36].

As noted, the small sample size makes achieving significant statistical results difficult, but other potential design related causes for negative results should be considered. For example, the controller, while able to operate in real time to provide suggestions to the participant as to which agent (automation or human) would be likely to better perform the task at a given time, was dependent on relatively brittle algorithms that used individual level data regarding expected driving behavior, but could not adapt to variability of individual participants. This inability to adapt the controller meant that the weight given to each of the components (behavioral data and the psychophysiological data) was fixed in the sensor fusion algorithm. In future experiments, we hope to develop approaches to dynamically assess the appropriate weight for each component in order to maximize overall performance.

Another potential challenge, and area of improvement, is that the algorithms used to predict whether an individual is likely to change control (switch from manual drive to autonomy drive or from autonomy drive to manual drive) were developed using group level data from a previous experiment [29] using a similar experimental setup. In that previous study, control toggles between manual drive and autonomy drives were relatively rare. Thus in this study, we were unable to collect enough individual specific training data to train our algorithms. We leveraged data from 15 participants in that previous study to build a group level model based on the recorded physiological, behavioral and environmental data. While this model has been shown effective at predicting control toggles in a manuscript currently under review elsewhere, it does not capture the individual differences that naturally occur between participants.

Additionally, it should be noted that we chose a visual actuator as opposed to one that uses a different modality. It is accepted that humans tend to ignore visual information over time as compared to when information or communication is presented using a different modality. This point reflects the importance of the choice of methods of communication within mixed-initiative teams in future studies. In this case, a visual actuator was chosen because of time and engineering constraints. Future studies may explore different option for multimodal interfaces to communicate with the human are likely to demonstrate improvements in performance.

In summary, given the limited data available, the inclusion of physiological information as a form of implicit communication seems to improve joint human-autonomy performance. Although there were no explicit measures of shared mental models or situation awareness, the literature suggests that communication between team members is critical to the development of these constructs within teams. We therefore suspect that improved team performance is reflective of successful development of shared mental models and situation awareness. Further analysis of this ongoing data collection will enable a more detailed interpretation of these data. Additionally, future work in this area will seek to expand these findings into other task and environment domains, as well as potentially exploring the mechanisms involved in improved team performance.

Acknowledgements. This research was supported the U.S. Office of the Secretary of Defense through the Autonomy Research Pilot Initiative (MIPR DWAM31168) as well as by an appointment to the U.S. Army Research Laboratory Postdoctoral Fellowship program administered by the Oak Ridge Associated Universities through a cooperative agreement with the U.S. Army Research Laboratory. We would also like to acknowledge the contribution provided by the Air Force Research Laboratory.

References

1. Teo, G., Reinerman-Jones, L.: Robot behavior for enhanced human performance and workload. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 117–128. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_12
2. Sheridan, T.B., Parasuraman, R.: Human-automation interaction. Rev. Hum. Factors Ergon. **1**(1), 89–129 (2005)
3. Parasuraman, R., Riley, V.: Humans and automation; use misuse, disuse, abuse. Hum. Factors **39**(2), 23 (1997)
4. Manzey, D., Bahner, J.E., Hueper, A.-D.: Misuse of automated aids in process control: complacency, automation bias and possible training interventions. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, no. 3, pp. 220–224. Sage Publications (2006)
5. McGuirl, J.M., Sarter, N.B.: Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. Hum. Factors **48**(4), 656–665 (2006)
6. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**(2), 381–394 (1995)
7. Kaber, D.B., Endsley, M.R.: Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. Process Saf. Prog. **16**(3), 126–131 (1997)
8. Cummings, M.L., Mastracchio, C., Thornburg, K.M., Mkrtchyan, A.: Boredom and distraction in multiple unmanned vehicle supervisory control. Interact. Comput. **25**(1), 34–47 (2013)
9. David, R.A., Nielsen, P.: Defense science board summer study on autonomy. Defense Science Board Washington United States (2016)
10. Schaefer, K.E., Straub, E.R., Chen, J.Y., Putney, J., Evans, A.: Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cogn. Syst. Res. **46**, 26–39 (2017)
11. Lackey, S., Barber, D., Reinerman, L., Badler, N.I., Hudson, I.: Defining next-generation multi-modal communication in human robot interaction. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, no. 1, pp. 461–464. SAGE Publications Sage CA, Los Angeles (2011)
12. Phillips, E., Ososky, S., Grove, J., Jentsch, F.: From tools to teammates: toward the development of appropriate mental models for intelligent robots. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 55, no. 1, pp. 1491–1495. SAGE Publications Sage CA, Los Angeles (2011)
13. Ososky, S., et al.: The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In: Proceedings of SPIE, vol. 8387, p. 838710 (2012)
14. Hill, S.G., Barber, D., Evans III, A.W.: Achieving the vision of effective Soldier-robot teaming: Recent work in multimodal communication. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, pp. 177–178. ACM (2015)

15. Kruijff, G.J.M., Janíček, M., Keshavdas, S., Larochelle, B., Zender, H., Smets, N.J.J.M., Mioch, T., Neerincx, M.A., Diggelen, J.V., Colas, F., Liu, M., Pomerleau, F., Siegwart, R., Hlaváč, V., Svoboda, T., Petříček, T., Reinstein, M., Zimmermann, K., Pirri, F., Gianni, M., Papadakis, P., Sinha, A., Balmer, P., Tomatis, N., Worst, R., Linder, T., Surmann, H., Tretyakov, V., Corrao, S., Pratzler-Wanczura, S., Sulk, M.: Experience in System Design for Human-Robot Teaming in Urban Search and Rescue. In: Yoshida, K., Tadokoro, S. (eds.) Field and Service Robotics. STAR, vol. 92, pp. 111–125. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-40686-7_8
16. Klien, G., Woods, D., Bradshaw, J., Hoffman, R., Feltovich, P.: Ten challenges for making automation a team player in joint human agent activity. IEEE Intell. Syst. **19**(6), 4 (2004)
17. Christoffersen, K., Woods, D.D.: How to make automated systems team players. In: Advances in human performance and cognitive engineering research. Emerald Group Publishing Limited, pp. 1–12 (2002)
18. Barber, D., Lackey, S., Reinerman-Jones, L., Hudson, I.: Visual and tactile interfaces for bi-directional human robot communication. In: SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, pp. 87410U (2013)
19. Oh, J., Howard, T.M., Walter, M.R., Barber, D., Zhu, M., Park, S., Suppe, A., Navarro-Serment, L., Duvallet, F., Boularias, A., Romero, O., Vinokurov, J., Keegan, T., Dean, R., Lennon, C., Bodt, B., Childers, M., Shi, J., Daniilidis, K., Roy, N., Lebiere, C., Hebert, M., Stentz, A.: Integrated intelligence for human-robot teams. In: Kulić, D., Nakamura, Y., Khatib, O., Venture, G. (eds.) ISER 2016. SPAR, vol. 1, pp. 309–322. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50115-4_28
20. Barber, D.J., Abich IV, J., Phillips, E., Talone, A.B., Jentsch, F., Hill, S.G.: Field assessment of multimodal communication for dismounted human-robot teams. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 59, no. 1, pp. 921–925. SAGE Publications Sage CA, Los Angeles (2015)
21. Kaupp, T., Makarenko, A., Durrant-Whyte, H.: Human–robot communication for collaborative decision making—A probabilistic approach. Robot. Auton. Syst. **58**(5), 444–456 (2010)
22. Schaefer, K.E., Evans, A., Hill, S.G.: Command and control in network-centric operations: trust and robot autonomy. In: 20th International Command and Control Research and Technology Symposium, Annapolis, MD (2015)
23. Kulic, D., Croft, E.A.: Affective state estimation for human–robot interaction. IEEE Trans. Robot. **23**(5), 991–1000 (2007)
24. Kulic, D., Croft, E.: Anxiety detection during human-robot interaction. In: 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2005, pp. 616–621. IEEE (2005)
25. Kulic, D., Croft, E.: Estimating robot induced affective state using hidden markov models. In: 2006 The 15th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2006, pp. 257–262. IEEE (2006)
26. Rani, P., Sarkar, N., Smith, C.A., Kirby, L.D.: Anxiety detecting robotic system–towards implicit human-robot collaboration. Robotica **22**(1), 85–95 (2004)
27. Dzindolet, M., Peterson, S., Pomranky, T., Beck, P.: The role of trust in automation reliance. Int. J. Hum.-Comput. Stud. **58**(6), 21 (2003)
28. Muir, B.M., Moray, N.: Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics **39**(3), 31 (1996)

29. Drnec, K., Metcalfe, J.S.: Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In: Schmorrow, D.D.D., Fidopiastis, C.M.M. (eds.) AC 2016. LNCS (LNAI), vol. 9744, pp. 157–167. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39952-2_16
30. Metcalfe, J.S., et al.: Building a framework to manage trust in automation. In: Micro-and Nanotechnology Sensors, Systems, and Applications IX. Vol. 10194. International Society for Optics and Photonics (2017)
31. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory—versions 4a and 54. Berkeley, CA. University of California, Berkeley, Institute of Personality and Social Research (1991)
32. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
33. Lee, J.D., Moray, N.: Trust, self-confidence, and operators' adaptation to automation. Int. J. Hum.-Comput. Stud. **40**(1), 153–184 (1994)
34. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics **35**(10), 1243–1270 (1992)
35. Carmody, M.A., Gluckman, J.P.: Task specific effects of automation and automation failure on performance, workload and situational awareness. In: Proceedings of the Seventh International Symposium on Aviation Psychology, vol. 1, pp. 167–171 (1993)
36. Ma, R., Kaber, D.B.: Situation awareness and workload in driving while using adaptive cruise control and a cell phone. Int. J. Ind. Ergon. **35**(10), 939–953 (2005)

Extending Embodied Interactions in Mixed Reality Environments

Mohamed Handosa, Hendrik Schulze, Denis Gračanin(✉), Matthew Tucker, and Mark Manuel

Department of Computer Science, Virginia Tech, Blacksburg, VA 24060, USA
gracanin@vt.edu

Abstract. The recent advances in mixed reality (MR) technologies provide a great opportunity to support deployment and use of MR applications for training and education. Users can interact with virtual objects that can help them be more engaged and acquire more information compared to the more traditional approaches. MR devices, such as the Microsoft HoloLens device, use spatial mapping to place virtual objects in the surrounding space and support embodied interaction with those objects. However, some applications may require an extended range of embodied interactions that are beyond the capabilities of the MR device. For instance, interaction with virtual objects using arms, legs, and body almost the same way we interact with physical objects. We describe an approach to extend the functionality of Microsoft HoloLens to support an extended range of embodied interactions in an MR space by using the Microsoft Kinect V2 sensor device. Based on that approach, we developed a system that maps the captured skeletal data from the Kinect device to the HoloLens device coordinate system. We have measured the overall delay of the developed system to evaluate its effect on application responsiveness. The described system is currently being used for the development of a HoloLens application for nurse aide certification in the Commonwealth of Virginia.

Keywords: Mixed reality · User gestures · Tracking

1 Introduction

Mixed Reality (MR) technologies allow for providing users with an environment that blends the physical surroundings with virtual objects. In order to support user interaction in such environment, there should be means to capture user inputs. Different MR devices tend to support user interaction in a variety of ways, where a user may provide input using physical controls, voice commands, and/or gestures. For instance, the ODG MR-glasses [22] has several input capabilities including on-device buttons and trackpad, a Wireless Finger Controller (WFC) with motion/gesture functionality, and a Wireless Bluetooth Keyboard with multifunction command keys. Another example is the Microsoft HoloLens

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 314–327, 2018.
https://doi.org/10.1007/978-3-319-91581-4_23

device [20], which uses spatial mapping to place virtual objects in the surrounding space and supports interaction with those objects through voice commands or gaze and air-tap. In gaze and air-tap interaction, the user gazes at the object of interest before making the air-tap gesture to trigger an action. Alternatively, the user may trigger the action by pressing a single-buttoned Bluetooth device, namely clicker.

The HoloLens' support for gesture-based input is very limited as it can recognize only two predefined gestures; air-tap and bloom. In several interaction scenarios, users might prefer to interact with virtual objects in a more engaging manner using their limbs almost the same way they interact with their physical counterparts. Unfortunately, this kind of embodied interaction is not supported by the current MR devices. However, support for user interaction in MR environments can be extended with the aid of other devices. Compared to the use of gaze and air-tap or the use of voice commands to interact with virtual objects, the support for embodied interaction can provide a more natural way to interact with the surroundings and allows for developing a rich user interface.

The remainder of the paper is organized as follow. Section 2 provides the related work. Section 3 explores the limitations of MR devices constraining their ability to support user interaction in MR environments. In Sect. 4, we present an approach to extend user interaction in MR environments. Based on that approach, we developed a system that integrates Microsoft HoloLens and Kinect devices. Section 5 demonstrates a case study, where the developed system enables user interaction in an MR space using different body joints. System performance (latency) is discussed in Sect. 6 while Sect. 7 concludes the paper.

2 Related Work

Virtual reality (VR) and Augmented Reality (AR)/MR technologies [5] technologies/applications are becoming more affordable and thus more accessible to general users. While there are over forty years of research in this area, we still need more findings to better understand challenges when it comes to developing MR applications. For example, in the education domain, MR [15]/AR [2], and VR [6] applications have been used mostly in higher education (science, humanities, and art), unlike vocation education. In the medical/health domain, MR has been mostly used for medical treatment, surgery, rehabilitation, education, and training, but not universally to other medical fields [7].

Some of the challenges in creating AR systems were explored by Dunleavy et al. [10]. They studied the limitation and affordances of an MR system. The results showed that although using MR system could significantly increase student engagement, there are still hardware and software issues. The advantages of MR include learning gains, motivation, interaction, and collaboration. Better learning performance, motivation, and engagement demonstrate the effectiveness of MR [2]. However, there is a need for longitudinal studies to study the evolution of knowledge and skills over time and to inform about the suitability of MR for supporting significant learning.

While being less immersive may be an inherent problem for MR technology, meanwhile it also proposes an interesting question for how we can expand the application scope for full utilization of this technology. One of the main challenges of MR is the limited field of view. As such, how to visualize large chunks of (or big) data is questionable [28]. With technology progressing forward, it is expected that the field of view can be enlarged even beyond of the human field of view in the near future [27,32] so this roadblock can be alleviated.

Another challenge of MR is that while it offers more naturalist interaction experiences [24], it is less immersive compared to the virtual reality technology. Hence, users can be distracted by environmental and other related factors leading to adverse impacts on usability [18]. This implies that MR is not a universal suitable technology for every different application.

MR systems inherently depend on the surrounding space to support user interactions. Our cognitive processes are dependent on how our body interacts with the world (affordances) [16,17] and how we off-load cognitive work onto our physical surrounding (embodied cognition) [30].

Embodied interactions demonstrate the importance of the body's interactions with the physical world. Interaction of our body and the surrounding physical world affect our cognitive processes and embodied cognition [31]. Embodiment cognition leverages the notion of affordances, potential interactions with the environment, to support cognitive processes [8]. Embodied interaction [8] and embodied user interfaces [11] lead towards invisible user interfaces and move the computation from desktop computers to physical space and place [9].

MR can also be used in a collaborative setting. Billinghurst and Kato explored the notion of functional and cognitive seams in collaborative MR systems [4] and reviewed MR techniques for developing collaborative interfaces. These results are reflected in collaborative and standalone MR applications [12]. Sharing gaze, emotions, and physiological cues can enhance collaboration in MR [25] and affect education outcomes [21] and collaboration [26]. The appearance of avatars/virtual agents can affect critical emotional reactions in interpersonal training scenarios as well as users' perceptions of personality and social characteristic [29].

3 Problem Definition

The main goal of MR is to enrich the actual physical environment with digital (virtual) entities. To achieve full immersion, the MR environment should react to user's behavior appropriately and the interaction should be as natural and intuitive as possible. The spatial awareness of an MR device, such as the Microsoft HoloLens device [20], allows a great degree of freedom regarding recognition, movement, and exploration of confined spaces and physical objects, enriched with virtual objects. However, the interaction capabilities of the HoloLens are limited by multiple factors which play an important role regarding the natural feeling to an immersive environment.

The interaction concept is based on voice commands, gaze tracking, hand tracking, and hand gestures. This concept yields the limitations of the device.

We focus on the limitations induced by gaze tracking, hand gestures, and hand tracking. For gaze tracking, the HoloLens uses its orientation to indicate its user's gazing direction. This assumption is not always true as a user may gaze at different directions while maintaining the same head orientation. Therefore, a virtual cursor is usually utilized to help the user perceive the gazing direction assumed by the HoloLens. Adding such extra virtual objects to an MR scene may not be the best way to support natural and intuitive interaction.

For hand gestures, the HoloLens supports only two core gestures, namely bloom and air-tap. The bloom gesture is reserved by the system to perform predefined special actions, such as showing/hiding the start menu and exiting from a running application. This limits the supported gestures when interacting with a HoloLens application to one gesture only, the air-tap, which is a transition between two recognizable hand states, ready and press as shown in Fig. 1.

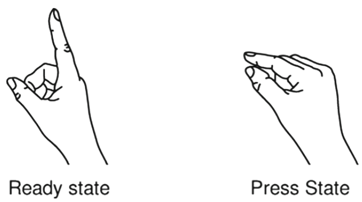


Fig. 1. Air-tap gesture, a switch from the ready to the press state [19].

For hand tracking, the HoloLens can track a hand position only if it is in the ready state (see Fig. 1), a closed fist with the index finger pointing up. Consequently, a user must maintain the ready hand state to enable hand tracking, which might be inconvenient, especially for long interaction scenarios. Moreover, the HoloLens cannot discriminate between left and right hands. In fact, the HoloLens treats a tracked hand as a disjoint object floating in space with no information about its side nor whether it belongs to the user or not. Consequently, the HoloLens may track the hand of a person other than the user and trigger actions accordingly, which can cause interaction conflicts in collaborative spaces with multiple users wearing HoloLens devices.

Both gesture recognition and hand tracking require the user's hand to be within the HoloLens' field of view. Additionally, gaze tracking follows the HoloLens' orientation rather than the actual gaze direction of the user, requiring the user to adjust the head orientation towards the object of interest rather than simply gazing at it. Both of these preconditions add a limitation to the possible *space of interaction* and create the need for not necessarily natural behavior patterns in order to interact with objects in a given environment. The lack of custom gestures recognition and full-body tracking (or at least discriminating left and right hands) limits the possible *range of interaction*. Natural interaction patterns such as interaction with both hands at the same time or interaction with multiple objects simultaneously are not (or only to a certain degree) possible.

Furthermore, the requirement of using unnatural hand gesture (ready state) in order to activate hand tracking and interact with the surroundings can greatly affect the *quality of interaction* as it interrupts the immersive experience and poses a difficulty to overcome, especially for users which are inexperienced with the usage of these gestures.

Integrating additional tracking devices, such as the Kinect V2 device, to observe the HoloLens user enables full-body tracking and identification of different body parts. Utilizing such information allows for developing more complex interaction schemes, involving multiple body parts and a higher level of detail for a broader range of recognizable gestures. For example, the recognition of the entire skeleton allows interaction with objects outside of the HoloLens' field of view and interaction with both hands at the same time. Moreover, interaction is not restricted to gestures performed with hands but can be extended to other body parts. The skeletal information, in combination with the spatial awareness of the HoloLens, allows for inferring contextual information from natural body movements.

4 Proposed Approach

We rely on tracking devices to capture the body movements and gestures of the user on behalf of the MR device. Providing such information allows MR applications to overcome the limitations of the MR devices and support rich interaction scenarios. Before MR devices can benefit from user tracking information, this information should be mapped from the tracking device coordinate system to the MR environment coordinate system. Registering two coordinate systems can be achieved by collecting a set of point pairs. Each pair consists of two corresponding points, one from each coordinate system. Once those points are collected, a registration algorithm can be applied to obtain a transformation matrix that maps a point from one world to the other. Several registration algorithms have been proposed such as the algorithm proposed by Besl and McKay [3] and the eight-point algorithm [14].

A system that integrates MR devices with user tracking devices is illustrated in Fig. 2. For each tracking device, there is a server application providing access to the tracking data provided by that device. The MR application should incorporate two modules, a client and a registration module. The client module is responsible for obtaining the tracking data from the server over the network interconnecting them while the registration module is responsible for mapping obtained data to the coordinate system of the MR device. Following this architecture, MR devices can obtain data from several tracking devices and a tracking device can provide data to several MR devices.

Some tracking devices can track several users simultaneously. Consequently, MR devices may receive several tracking data sets for several persons. In that case, the MR device may need to identify which data set belongs to the user. Given that MR devices are head-mounted devices, the current location of a device in the MR coordinate system gives a good indication of the current location of

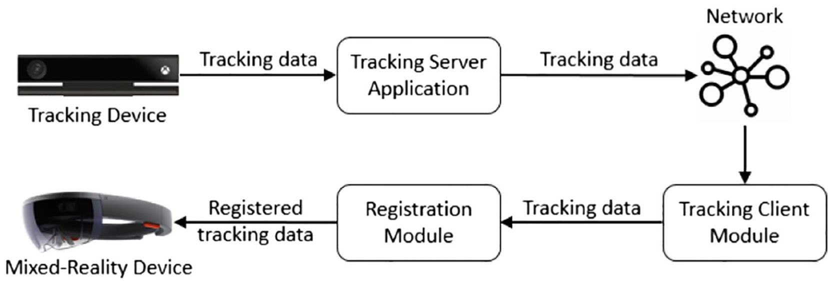


Fig. 2. Integrating MR devices and tracking devices.

the user's head. Comparing the device location with the registered tracking data sets can reveal which data set belongs to the user.

4.1 Implementation

Based on the proposed approach, we implemented a system that integrates HoloLens devices with Kinect devices. A Kinect server application tracks the user skeleton using the Kinect device. The HoloLens application obtains tracking data from the server through its Kinect client module before the registration module maps it to the HoloLens coordinate system using a transformation matrix. In order to obtain the transformation matrix, we have developed a four-step process (Fig. 3).

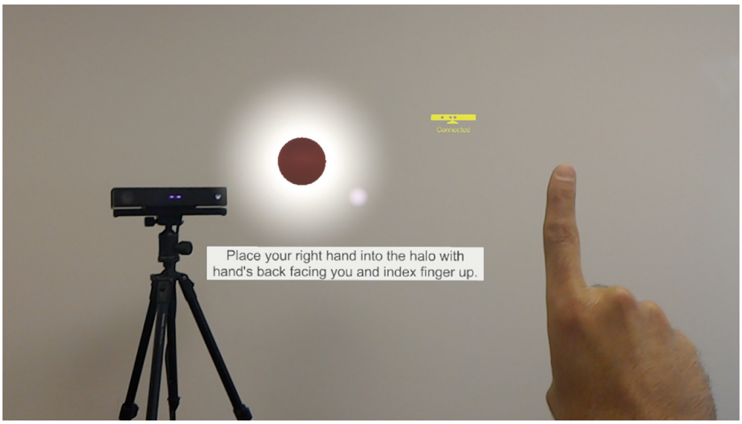


Fig. 3. Calibration process.

The goal of each step is to collect two corresponding points, one from the Kinect coordinate system and another from the HoloLens coordinate system. The four-point pairs are collected by asking the user to place a hand at four different positions in space that are indicated by virtual objects. Once the user's hand is in position, hand tracking information is collected from both Kinect and HoloLens to form a point-pair. Afterward, an algorithm is applied to obtain the transformation matrix and save it for later use. Figure 4 shows a HoloLens-rendered virtual skeleton aligned with the corresponding physical body.

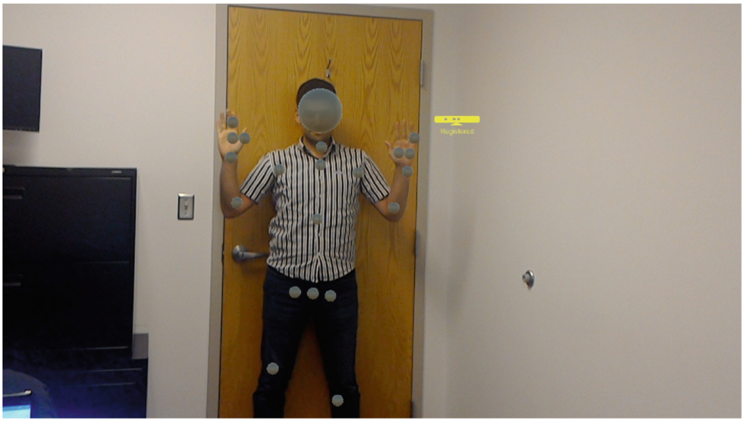


Fig. 4. A registered skeleton aligned with the corresponding physical body.

4.2 Network Infrastructure

In order to communicate tracking data from the tracking server to the tracking client, we have tested two communication models, direct and indirect (Fig. 5). For direct communication, we use the User Datagram Protocol (UDP). The server has a predefined listening port to which clients can send subscription requests. The server collects tracking data from the Kinect device before sending it to all subscribing clients. This communication model minimizes the communication delay. However, for multiple-Kinect setup, a HoloLens will need to communicate with multiple servers. Establishing several connections with different servers complicates the networking model and makes network troubleshooting more challenging.

In order to support multiple-Kinect/multiple-HoloLens setups while minimizing the communication model complexity, we use the Message Queue Telemetry Transport (MQTT) protocol [1,13]. MQTT is a publish/subscribe communication protocol that relies on a broker to support indirect communication between publishers and subscribers. Each Kinect server can publish tracking data to a specific topic on the MQTT broker. Unlike direct communication, a client will need to maintain only a single connection with the MQTT broker. The client can subscribe to one topic or more to receive tracking data from one or more Kinect servers. Although, indirect communication may introduce increased delays, it allows for relaxing the complexity of the communication model. Furthermore, additional data sources, including environmental and biometric sensors, can be added.

Fig. 5. Two alternative communication models: (a) UDP-based direct communication and (b) MQTT-based indirect communication.

5 Case Study

In order to evaluate our approach in a non-lab environment, we used it in the development of a HoloLens application for Nurse Aide skills training [23]. The goal of the application is to augment the student's experience in classroom settings and to provide a rich set of educational contents in an MR environment. Students need to learn a set of skills with quite specific steps in a certain order. This requires not only the theoretical knowledge of how a specific skill needs to be done but also practical application in order to manifest the exact workflow required. With limited space and limited availability of hospital equipment in schools, the number of workstations to actually practice the skills are limited as well.

We developed a HoloLens application which recreates the scenery of hospital room. Within this virtual hospital room (Fig. 6a) are the required objects and props to perform the skills in a 'close to reality' environment. Figure 6b demonstrates an embodied interaction with digital entities, a denture, and a toothbrush. Our application guides the student through the steps of a skill and requires the student to perform specific and detailed interactions within the MR environment in order to proceed to the next step of a skill.

Fig. 6. (a) The virtual hospital room. (b) Demonstration of using both hands simultaneously for brushing a denture.

Almost all skills require at some point a more detailed user tracking than the HoloLens alone can provide. For example, a crucial part of proper hand washing requires the student to keep the hands and forearms at a downward angle to prevent 'contaminated' water to run down the arms. With the HoloLens alone, there is no possibility to check this condition. Another example is denture brushing where a student should hold a denture in one hand and a toothbrush in the other hand. With HoloLens alone, enabling hand tracking will require the student to maintain both hands in the ready state (Fig. 1) and within the HoloLens field of view, resulting in constrained and unnatural interaction. However, with the additional data about the entire user skeleton provided by the Kinect, we were able to achieve a level of detail and precision to track the user's actions sufficiently.

The application relies on HoloLens-based gesture recognition to support navigation through its menus and to adjust different application settings. However, once a given skill training is started, the application relies on Kinect data to support user interaction. As most of the students had little to no experience with MR devices and MR environments in general, it took some 'warm-up' time for the students to get used to the new experience and to move around and interact with virtual objects comfortably. While the gestures recognized using Kinect data were easy to learn, the original HoloLens gestures (air-tap) required some time to learn. After the initial learning phase, students were able to complete the skill training without or with minimal further guidance. Integrating the Kinect device into the system allowed the students to avoid the difficulties associated with the use of unnatural gestures, which helps reducing the threshold for users to comfortably interact with the virtual objects.

6 Results

Users should receive instant feedback as they interact with objects in an MR space. A noticeable delay in response to user commands can degrade the user experience dramatically. Therefore, the user's skeletal information should be delivered to the HoloLens with minimal latency to ensure the responsiveness of the system. The responsiveness is determined by the delay (latency) between the time at which the user makes a given gesture/move and the time at which the user receives the corresponding feedback through the MR device.

For the purpose of estimating the overall latency, we have captured multiple MR video recordings of user gestures, specifically the closed hand gesture. Exploring the frames of the captured videos revealed that it takes at most four frames for the HoloLens to provide a feedback after the user gesture takes place. Figure 7 shows six consecutive frames from a captured MR video (30 frames per second). The user starts with an open hand and the HoloLens displays a red box indicating that the hand state is open (Fig. 7a). The user starts to close the hand but it is not closed yet (Fig. 7b). The user's hand is closed (Fig. 7c). The user is waiting for the feedback (Fig. 7d, e). The user receives a visual feedback, where a green box is shown (Fig. 7f).

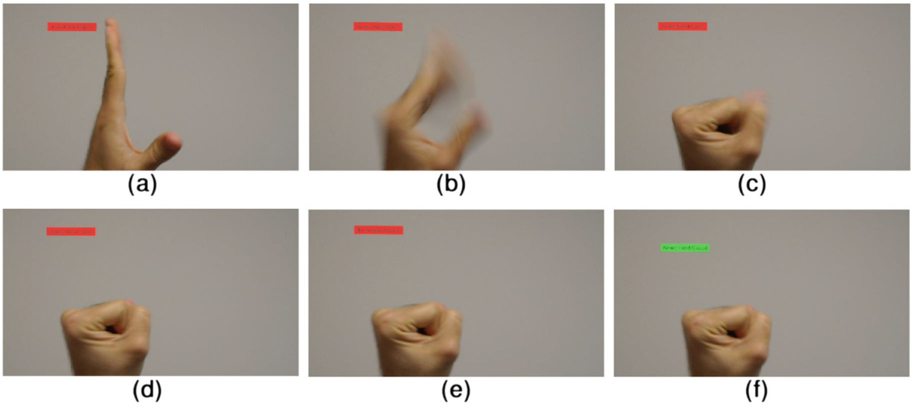


Fig. 7. Video recording consecutive frames for detecting a closed hand gesture using Kinect and providing feedback through HoloLens.

Assuming that frame *(b)* was captured at time 0 and frame *(c)* was captured at time T, then the user gesture takes place at time t_1, $0 < t_1 \leq T$. Similarly, if frame *(e)* was captured at time $3T$ and frame *(f)* was captured at time $4T$, then the feedback has occurred at time t_2, $3T < t_2 \leq 4T$. Consequently, the delay d is $2T < d < 4T$. The video frame rate is 30 frames per second and hence $T = 1/30\,\text{s}$ or 33.33 ms. Therefore, the total system delay ranges between 66 and 134 ms. This estimated latency is caused by the processing and communication steps that take place between a change in user skeleton state and providing the corresponding feedback.

The overall latency $d \geq d_1 + d_2 + d_3 + d_4$, where d_1 is the time it takes the Kinect device to capture a frame and send its data to the workstation; d_2 is the time it takes the workstation to extract skeleton information from the received frame data producing a skeleton information message; d_3 is the time needed to send the skeleton information message from the workstation to the HoloLens; and d_4 is the time it takes the HoloLens to provide a feedback based on the received skeleton information. Delays d_1 and d_4 are device specific and beyond our control. The measured average values of d_2 and d_3 (using UDP-based direct communication) are approximately 0.157 and 0.476 ms, respectively. Compared to the overall latency, both d_2 and d_3 are negligible.

The HoloLens can recognize the press gesture (Fig. 1). On the other hand, the Kinect device can recognize a closed hand gesture. Benefiting from the similarity between the closed hand gesture and the press gesture, we were able to measure the relative latency of the Kinect-based gesture recognition using the HoloLens-based gesture recognition as a reference point. Although the Kinect-based recognition involves several processing and communication steps, results have shown that its performance is comparable to that of the built-in HoloLens recognition. In fact, we noticed that the Kinect-based recognition can often perform faster than the built-in HoloLens gesture recognition. Figure 8 shows a

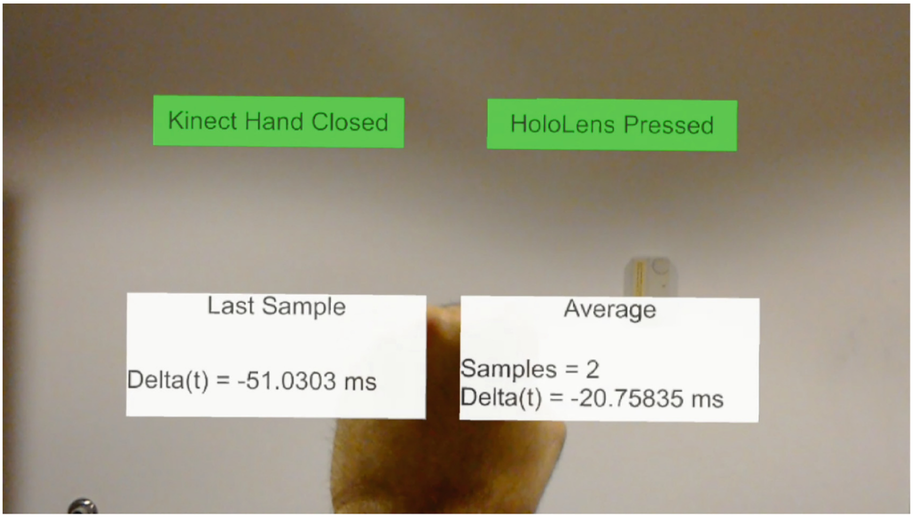


Fig. 8. A closed hand (or press) gesture recognized by both Kinect and HoloLens.

closed hand (or press) gesture, where the Kinect-based recognition outperformed the HoloLens-based recognition by approximately 51 ms.

Using MQTT-based indirect communication can simplify the communication model and make network troubleshooting less challenging, especially for multiple-Kinect/multiple-HoloLens setups. However, a significant increase in communication latency can degrade the user experience. The MQTT protocol supports three Quality of Service (QoS) levels. The message delivery for QoS-0, QoS-1, and QoS2 are at-most-once, at-least-once, and exactly-once, respectively.

Test results have shown that the average delay for the QoS-0, QoS-1, and QoS-2 are approximately 2.743, 28.492, and 36.047 ms, respectively. Although QoS-0 provides the smallest average delay, it allows for message dropping. However, this should not be a problem for applications that are interested in receiving the most recent tracking sample rather than receiving every tracking sample. Compared to the UDP-based direct communication average, the MQTT-based indirect communication with QoS-0 does not introduce a significant increase considering the overall delay of the system (Table 1).

Table 1. Average delay (latency) of UDP-based and MQTT-based communication.

	UDP	MQTT		
		QoS-0	QoS-1	QoS-2
Average delay (milliseconds)	0.476	2.743	28.492	36.047

7 Conclusion

MR devices provide an affordable opportunity to develop immersive applications. However, their limited input capabilities constrains the possible interactions. We presented an approach to overcome this limitation by integrating MR devices with tracking devices. An MR application can rely on the user tracking information to extend its ability to capture user inputs and to support interaction scenarios that were not possible before. Based on the proposed approach, we developed a system to integrate Microsoft HoloLens and Kinect devices. We presented a case study, where we utilized the developed system to support user interaction in an MR space using different body joints. The system allows users to interact with virtual objects and receive the corresponding feedback within 66 to 134 ms.

In order to support communication between Kinect devices and HoloLens devices, we have tested two communication models, UDP-based direct communication and MQTT-based indirect communication. Test results have shown that UDP-based communication introduces less average delay (0.476 ms) compared to MQTT-based communication with QoS-0 (2.743 ms). Although the MQTT-based communication introduces a slightly larger delay, it helps relaxing networking and communication complexities, especially for multiple-Kinect/multiple-HoloLens setups.

Current technical constraints and hardware limitations make a comprehensive solution that combines the multitude of required sensors with VR and MR devices difficult to achieve. However, there are many opportunities for combining multiple conventional and unconventional data sources (not only tracking devices) in a comprehensive framework. When the data sources provide overlapping coverage of the MR space, they can be used for internal data alignment, error correction, and increased accuracy.

References

1. Arlitt, M., Marwah, M., Bellala, G., Shah, A., Healey, J., Vandiver, B.: MQTT version 3.1.1 plus errata 01. Standard, OASIS, 10 December 2015
2. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented reality trends in education: a systematic review of research and applications. J. Educ. Technol. Soc. **17**(4), 133–149 (2014)
3. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992)
4. Billinghurst, M., Kato, H.: Collaborative mixed reality. In: Proceedings of the First International Symposium on Mixed Reality (ISMR 1999), pp. 261–284. Springer, Berlin (1999)
5. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. A K Peters, Wellesley (2005)

6. Calvo, O., Molina, J.M., Patricio, M.A., Berlanga, A.: A propose architecture for situated multi-agent systems and virtual simulated environments applied to educational immersive experiences. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H. (eds.) IWINAC 2017. LNCS, vol. 10338, pp. 413–423. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59773-7_42
7. Chen, L., Day, T.W., Tang, W., John, N.W.: Recent developments and future challenges in medical mixed reality. In: Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), pp. 123–135, October 2017
8. Dourish, P.: Where the Action is: The Foundations of Embodied Interaction. The MIT Press, Cambridge (2001)
9. Dourish, P.: Re-space-ing place: "place" and "space" ten years on. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW 2006), pp. 299–308. ACM, New York, 4–8 November 2006
10. Dunleavy, M., Dede, C., Mitchell, R.: Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. J. Sci. Educ. Technol. **18**(1), 7–22 (2009)
11. Fishkin, K.P., Moran, T.P., Harrison, B.L.: Embodied user interfaces: towards invisible user interfaces. In: Proceedings of the IFIP TC2/TC13 WG2.7/WG13.4 Seventh Working Conference on Engineering for Human-Computer Interaction, pp. 1–18. Kluwer, B.V., Deventer (1998)
12. Gračanin, D., Eck II, T., Silverman, R., Heivilin, A., Meacham, S.: An approach to embodied interactive visual steering: bridging simulated and real worlds. In: Proceedings of the 2014 Winter Simulation Conference (WSC), pp. 4073–4074, December 2014
13. Gračanin, D., Handosa, M., Elmongui, H., Matković, K.: An approach to user interactions with IoT-enabled spaces. In: Proceedings of the 14th International Conference on Telecommunications (ConTEL), pp. 139–146, 28–30 June 2017
14. Hartley, R.I.: In defense of the eight-point algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **19**(6), 580–593 (1997)
15. HoloGroup: Education in mixed reality. http://holo.group. Accessed 6 Jan 2018
16. Kirsh, D.: Embodied cognition and the magical future of interaction design. ACM Trans. Comput.-Hum. Interact. **20**(1), 3:1–3:30 (2013)
17. Klemmer, S.R., Hartmann, B., Takayama, L.: How bodies matter: five themes for interaction design. In: Proceedings of the 6th ACM conference on Designing Interactive systems (DIS 2006), pp. 140–149. ACM Press, New York (2006)
18. Lebeck, K., Kohno, T., Roesner, F.: How to safely augment reality: challenges and directions. In: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications, pp. 45–50. ACM, New York (2016)
19. Microsoft Corporation: How to perform an air tap. https://developer.microsoft.com/en-us/windows/mixed-reality/gestures. Accessed 2 Mar 2018
20. Microsoft Corporation: Microsoft HoloLens. https://www.microsoft.com/microsoft-hololens/. Accessed 2 Mar 2018
21. Noteborn, G., Carbonell, K.B., Dailey-Hebert, A., Gijselaers, W.: The role of emotions and task significance in virtual education. Internet High. Educ. **15**(3), 176–183 (2012)
22. ODG: Osterhout Design Group. https://www.osterhoutgroup.com/. Accessed 2 Mar 2018
23. Pearson: Virginia Nurse Aide: written (or oral) examination & skills evaluation, March 2017

24. Peters, E., Heijligers, B., de Kievith, J., Razafindrakoto, X., van Oosterhout, R., Santos, C., Mayer, I., Louwerse, M.: Design for collaboration in mixed reality: technical challenges and solutions. In: Proceedings of the 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), pp. 1–7, September 2016
25. Piumsomboon, T., Lee, Y., Lee, G.A., Dey, A., Billinghurst, M.: Empathic mixed reality: sharing what you feel and interacting with what you see. In: Proceedings of the 2017 International Symposium on Ubiquitous Virtual Reality (ISUVR), pp. 38–41, June 2017
26. Quarles, J., Lampotang, S., Fischler, I., Fishwick, P., Lok, B.: Collocated AAR: augmenting after action review with mixed reality. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR 2008), pp. 107–116, September 2008
27. Rakkolainen, I., Turk, M., Höllerer, T.: A compact, wide-FOV optical design for head-mounted displays. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, pp. 293–294. ACM, New York (2016)
28. Ren, D., Goldschwendt, T., Chang, Y., Höllerer, T.: Evaluating wide-field-of-view augmented reality with mixed reality simulation. In: Proceedings of the 2016 IEEE Virtual Reality Conference (VR 2016), pp. 93–102. IEEE (2016)
29. Volonte, M., Babu, S.V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S.B., Fasolino, T.: Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. IEEE Trans. Vis. Comput. Graph. **22**(4), 1326–1335 (2016)
30. Williams, A., Kabisch, E., Dourish, P.: From interaction to participation: configuring space through embodied interaction. In: Beigl, M., Intille, S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 287–304. Springer, Heidelberg (2005). https://doi.org/10.1007/11551201_17
31. Wilson, M.: Six views of embodied cognition. Psychono. Bull. Rev. **9**(4), 625–636 (2002)
32. Xiao, R., Benko, H.: Augmenting the field-of-view of head-mounted displays with sparse peripheral displays. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 1221–1232. ACM (2016)

Interaction of Distant and Local Users in a Collaborative Virtual Environment

Adrian H. Hoppe[1(✉)], Roland Reeb[1], Florian van de Camp[2], and Rainer Stiefelhagen[1]

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{adrian.hoppe,rainer.stiefelhagen}@kit.edu, roland.reeb@student.kit.edu
[2] Fraunhofer IOSB, Karlsruhe, Germany
florian.vandecamp@iosb.fraunhofer.de

Abstract. Virtual Reality enables a new form of collaboration. It allows users to work together in the same virtual room regardless of their actual physical location. However, it is unclear which effect the physical location of the user has on task performance, the feeling of presence or immersion. We compared the collaboration of two users in the same local room and in remote rooms on the basis of a knowledge-transfer task. An instructor indicated different virtual objects using three different pointing gestures and a trainee selected the highlighted object. The results of a 28 participant user study show that the performance of the gestures in the local and remote setup is equal, according to NASA-TLX, rankings and time. Users feel equally co-present and tend to prefer the remote collaboration. The data presented in this paper shows that VR collaboration in a virtual room is independent from the physical location of the participants. This allows the development of VR applications without special consideration of the user's location. VR systems can use the advantages of a remote collaboration, like faster reaction times, no travel expenses and no user collision, or of local collaboration, e.g. direct contact between users.

1 Introduction

Virtual Reality (VR) technology becomes more ubiquitous and a main focus is user collaboration in the Virtual Environment (VE). The collaborative VE [3] allows multiple users to analyze and discuss information as well as interact with the VE and each other. One key advantage of VR with regard to cooperation is that users can meet in an interactive VE without actually being at the same physical location. This eliminates travel time and expenses and saves users the effort to discuss complex information via telephone or video conference by providing a hands-on experience. One disadvantage of a remote interaction can be a delayed interaction due to network latencies. As stated, VR allows remote collaboration but it has yet to be investigated if there are any differences between interacting with other users at the same location (*local*) or at different physical locations (*remote*) (see Fig. 1). Hence, the impact of the user's physical location has to

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 328–337, 2018.
https://doi.org/10.1007/978-3-319-91581-4_24

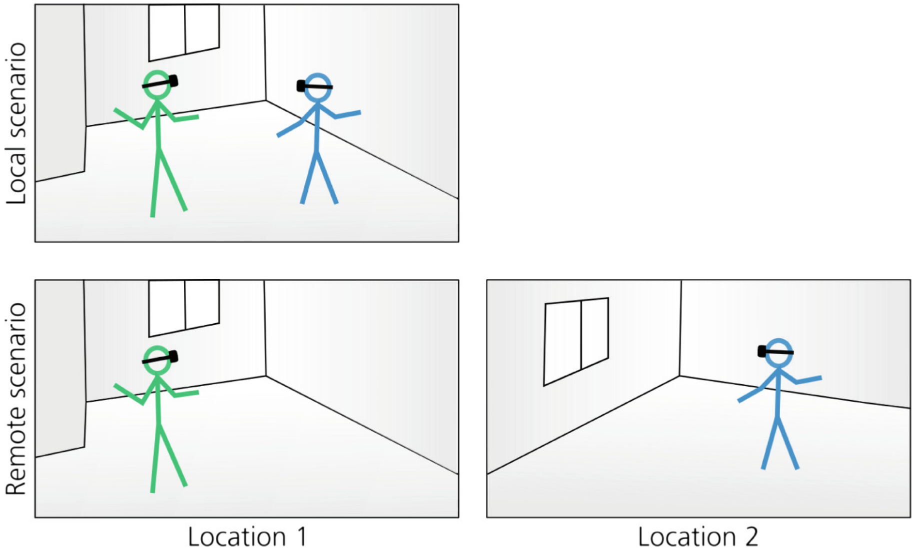


Fig. 1. In the local scenario both users are in the same physical room and the same virtual room. In the remote scenario, each user is in a separate room located in different buildings but both users are in the same virtual room.

be determined. An obvious advantage of a local collaboration is the possibility to physically interact with the other user, e.g. to exchange tools. Furthermore, users can directly communicate by speech without using any additional communication device. However, local collaboration also has some disadvantages. Local users in VR need to be aware of the location of the other user to avoid collisions with them. Individual locomotion of local users poses additional challenges: The physical location of another user and his or her virtual location can be different. In that case, the virtual avatar of the other users cannot be used for collision handling. Also, direct speech communication can be misleading, since the direction of the user's avatar differs from the direction of the user's voice. In conclusion, local and remote collaboration have advantages and disadvantages. This paper addresses the differences between both setups regarding task performance and user experience.

The paper is structured as follows. Section 2 gives an overview of related work. Section 3 presents a setup for local and remote collaboration in VR. The proposed setups are evaluated and compared and the results of a user study are presented in Sect. 4. Results are further discussed in Sect. 5, and Sect. 6 draws a conclusion.

2 Related Work

Salzman et al. [11] developed a cooperative system where two users can work together in VR in a local setup. To achieve this, they used two Head Mounted

Displays (HMD). The two users collaborated in an assembly task, where a windshield was inserted into a car. In the real world the windshield is a metal requisite which is tracked and virtually represented as a windshield. Users did not move in this setup, because the windshield was within reach of their start location. Furthermore this work focus on requisite-based interaction which is not possible in a remote setup. Beck et al. [2] developed a immersive VR system that allows group-to-group telepresence. A group can view the virtual world on a projection screen with shutter glasses. A control device in front of the screen allows group locomotion. The group is captured with a camera and displayed in the virtual world. Two hardware setups allow two groups to interact with each other in VR. Other systems that allow collaboration in VR are Studierstube [13] and the PIT [1]. Kranstedt et al. [7] investigate collaborative pointing. Two users stand on opposite sides of a table, which has different parts of a model plane on it. One user, the *description giver*, sees an exact virtual representation of the table. His or her hand is tracked. The hand is represented in VR with an additional laser pointer, which allows the description giver to point on different locations on the table. The other user, the *object identifier*, is not in VR. He or she sees the pointing gesture of the description giver and identifies the pointed-at object with a pointer. Results of this study do not contain user collaboration since only the geometric pointing behavior was researched.

3 Collaboration in VR with Local and Remote User Locations

In order to compare local and remote collaboration in VR we determined the requirements for a VR system. To allow users to view the VE and interact with it independently, we installed three HTC Vive HMDs in two different locations. For the local scenario, we set up two HTC Vive, connected to two computers, in the same room. Since the HMD cables of the users in the local scenario could lead to users falling and a wireless connection for the HMDs was not available, the cables to the hmd were suspended from the ceiling of the room. The length of the cable was adjusted with a retractable dog leash. However, the cable of the third Vive in the remote room lies on the floor. For the remote scenario, we set up the HTC Vives in two different rooms respectively. The two rooms are located in different buildings but connected by a 1Gbps Ethernet network.

A user is represented by an avatar which is aligned using the head and hand positions through inverse kinematics (see Fig. 2). The avatar is important for the feeling of co-presence [12]. The stylized representation has no significant difference to a human avatar [6] and avoids the uncanny valley [8]. Roth et al. [10] determined that non-realistic avatars handicap social interactions. However missing behavioral characteristics, like gaze or facial expressions, can be partially compensated by using other behavior channels, like gestures. They concluded that a mannequin is a universal representation of a human, which is easy to reproduce and animate.

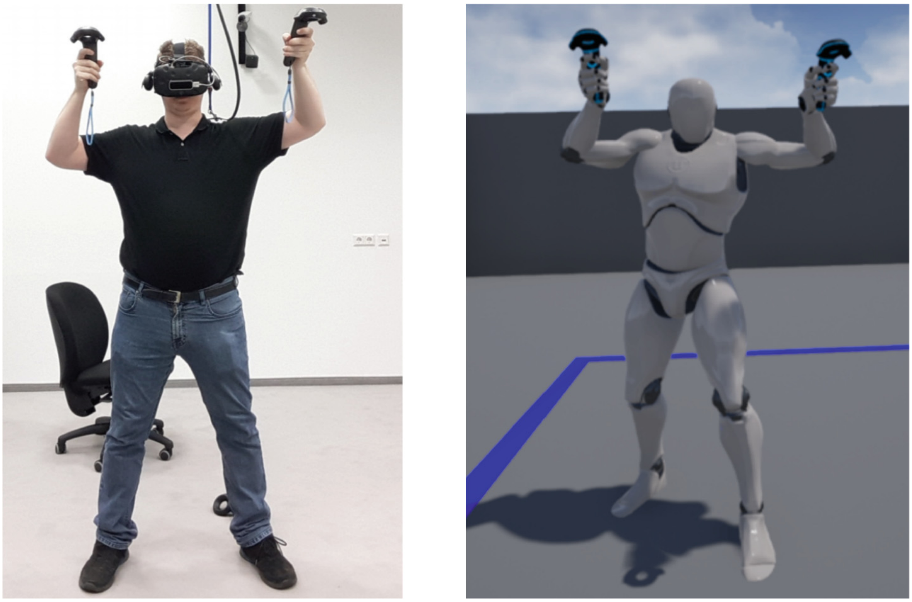


Fig. 2. Avatar representation of the user in VR.

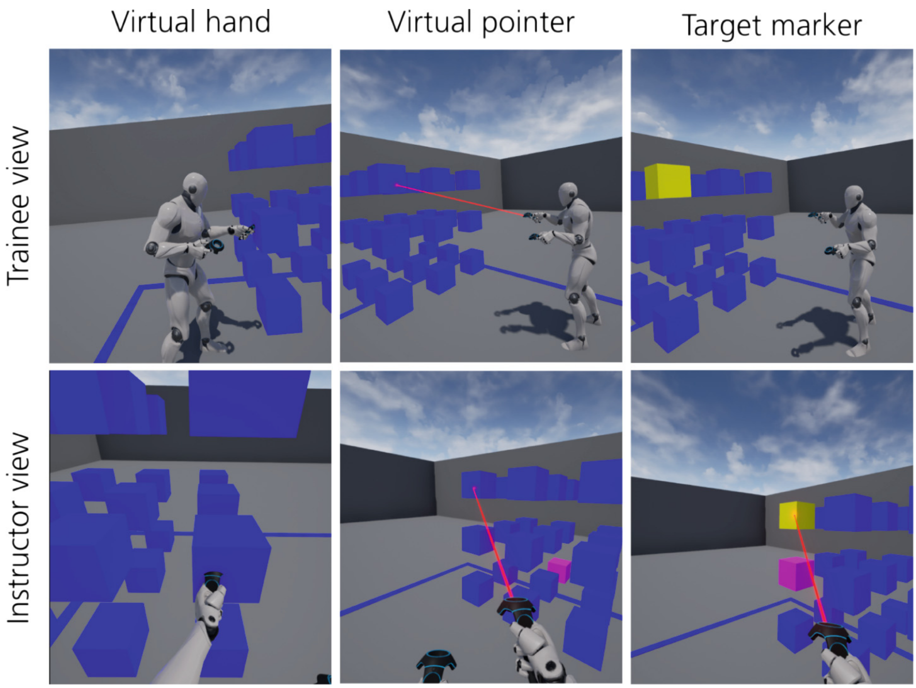


Fig. 3. All three pointing techniques from left to right with the trainee's view on top and the instructor's view on the bottom.

In both scenarios, users used a Logitech G930 headset to communicate via voice chat with each other. The direction of the audio signal of a speaking user is adjusted according to the location of his or her avatar.

To assure user collaboration we implemented a knowledge-transfer scenario where two users take different roles. One user, the *instructor*, highlights virtual objects for another user using a pointing gesture. The second user, the *trainee*, then needs to interact with the indicated object by selecting it using direct touch. To evaluate the effects of the two collaboration setups, three different pointing gestures were examined. The used pointing gestures are virtual hand, virtual pointer [9] and target marker (see Fig. 3). With target marker, the instructor has a virtual laser pointer attached to his or her hand. In addition to that the pointed-at object is highlighted. The trainee also sees the visual highlight but not the beam of the laser pointer, since it is not absolutely necessary. The virtual objects are represented by cubes, arranged in a grid of 3 × 3 × 3 as in [14]. To increase task difficulty the grid has a static and a rotating mode. In the rotating mode the whole grid rotates around two axes with different velocities.

4 Evaluation

We conducted a user study with 30 participants which performed tasks in pairs of two. Due to technical problems one team was excluded, so the evaluation is based on the remaining 28 participants. 19 participants are male and 9 are female. Their average age was 22.5 years. On a five-point Likert scale from 0 (none) to 4 (high) the participants rated their experience with computer games with Ø 3.29 ± 0.81 and with VR with Ø 1.11 ± 1.10.

Each pair performed the tasks in both the remote scenario as well as the local scenario. The roles of the users were switched when the scenario changed. Both users performed all three gestures in the role of the instructor. To minimize the effects of learning and fatigue in the evaluation, the order of the scenarios, roles and gestures was randomized.

A task consists of three training rounds (two with static grid, one with rotating grid) and six timed rounds (three with static grid, three with rotating grid). One round contained one indication of the instructor and the interaction of the trainee with the virtual object. The round starts with both users standing on designated start positions and ends with the selection of the virtual object by the trainee.

To compare the two scenarios, qualitative and quantitative data was collected. Participants were asked how pleasant the collaboration with the partner was. Users rated the collaboration on a scale from 0 (very unpleasant) to 4 (very pleasant) with Ø 3.93 ± 0.26. This value shows that the pairs could work well together and the results are not negatively affected by a user's refusal to cooperate. Furthermore users were asked if they experienced nausea to check if the collected data could be negatively influenced. Users reported almost no nausea with Ø 3.79 ± 0.42 with 0 being strong nausea and 4 no nausea.

Figure 4 shows the users preferences for the different pointing gestures sorted by the room setup. Participants were asked to rank the gestures from first place

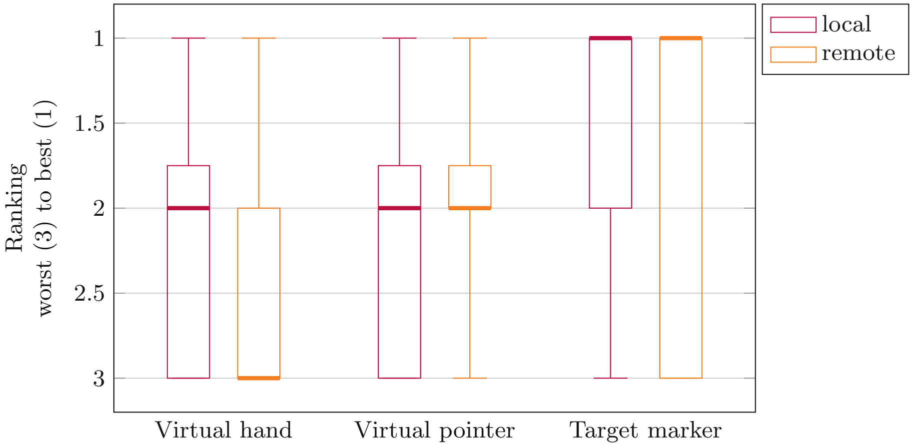


Fig. 4. Ranking of the pointing gestures with box-and-whisker plots performed by the instructor.

(1) to last place (3). Since no user performed the gestures in the role of the instructor in both local and remote setup the sample is independent an the Mann-Whitney-U-Test is used to check for significant differences. The median of 1 of target marker shows that this technique did perform well. Pairwise comparisons between the two room setups show that there are no significant differences for any of the pointing gesture ($p = 0.378$ for virtual hand, $p = 0.120$ for virtual pointer, $p = 1.000$ for target marker).

NASA-TLX rankings of the instructor show a low mental, physical and temporal demand with high performance and low effort and frustration levels (see Fig. 5). After applying an ANOVA, the results show no significant differences in a pairwise Dunn-Bonferroni test, except for the ranking of the physical demand of the gestures. The virtual hand interaction is more physically demanding with $p < 0.023$ in a pairwise comparison. The results of the NASA-TLX questionnaire for the trainee are similar to those of the instructor. However, no significant differences between the pointing gestures occur. This is as expected since the interaction of the trainee did not change, when the instructor changed to another pointing gesture.

Average interaction times per round are about four to five seconds, as shown in Fig. 6. The differences between the two setups are significant in the case of the virtual pointer ($p = 0.003$) and target marker ($p = 0.012$) according to a sign test. The effect size [5] can be described as medium ($r = 0.321$) and low ($r = 0.274$) respectively. Since the tasks are identical in both setups further investigations were performed. Both the time it took the instructor to point to the target object the first time and the number of correctly and incorrectly indicated virtual cubes are not significantly different. As a result, the speed of the trainee seems to be different in the local and remote setup.

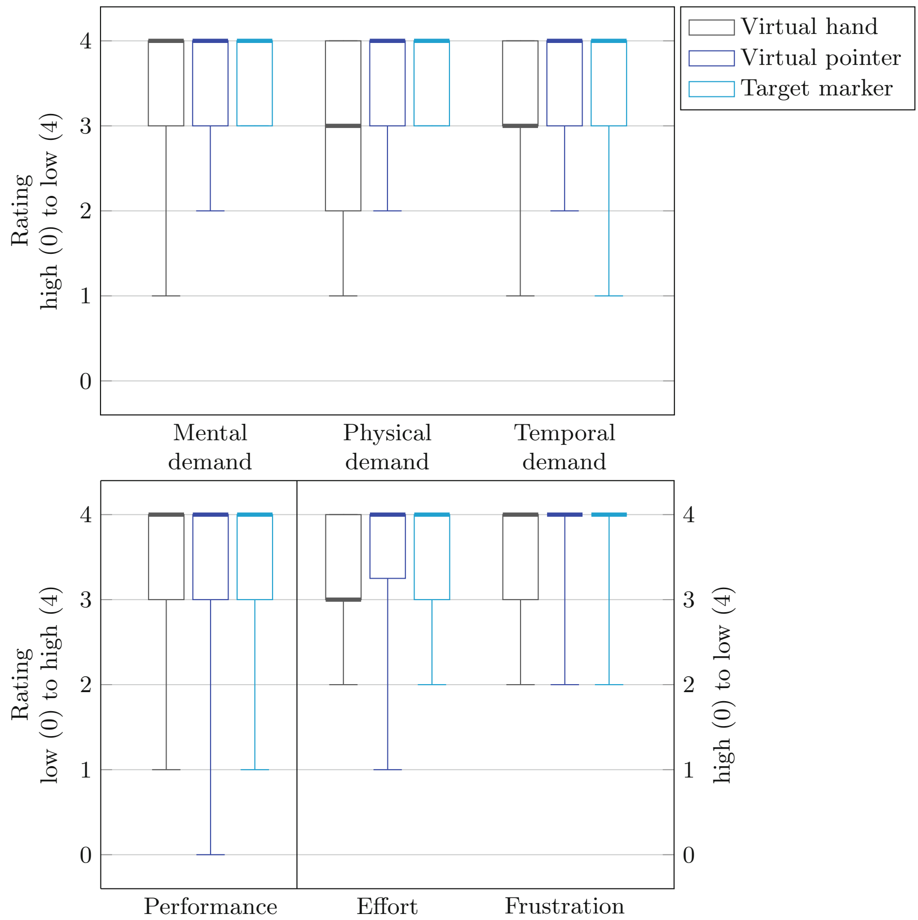


Fig. 5. NASA-TLX with box-and-whisker plots for pointing gestures performed by the instructor.

When asked 50% of the users did not prefer either one setup. 11% preferred the local interaction and 39% liked the remote interaction better. Ten out of eleven users explained their preference for the remote setup by saying that they did not need to worry about any collisions with the other party while working remotely. The other user was impressed by the capabilities of the collaboration via network. From the three users, who preferred the local setup, two said the collaboration is more realistic and one said that he did perceive the other user more as a human rather than a robot.

Users were asked how much they depended on speech communication while solving the tasks on a scale from 0 (not at all) to 4 (very much). A sign test for the ratings of Ø 0.43 ± 0.02 in the local setup and Ø 0.39 ± 0.02 in the remote setup shows no significant difference between the two scenarios ($p = 0.774$).

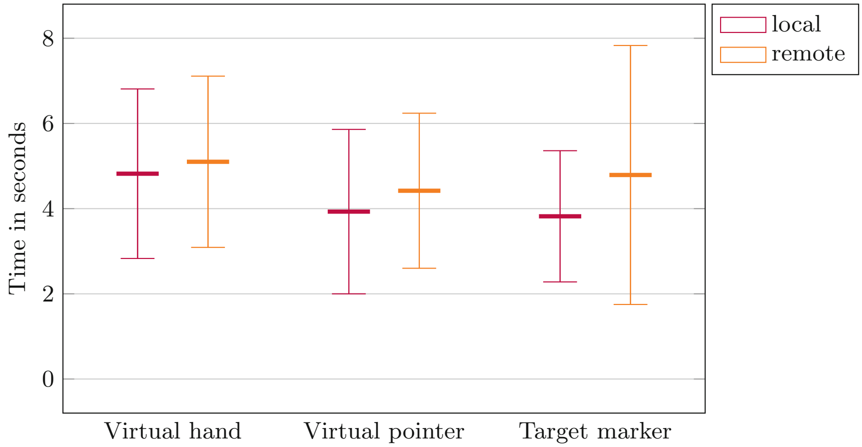


Fig. 6. Timings with average and standard deviation, sorted by local and remote setup.

Furthermore, the users rated the amount of co-presence they experienced with the other user, while performing the tasks of the user study. Co-presence was assessed on a scale from 0 (users feel like they are in different rooms) to 4 (users feel like they are in the same room). In the local setup users rated co-presence as $\varnothing\ 2.97 \pm 0.03$ and as $\varnothing\ 2.82 \pm 0.03$ in the remote setup. A sign test shows that the difference is not significant ($p = 0.092$).

5 Discussion

The results of the user study show that in general all gestures are suitable for the given task. No gesture outperformed any of the other gestures in all aspects. This result conforms with the conclusion of Bowman et al. [4] that all interaction techniques in VR have their strength and weaknesses and that there is no best technique.

A comparison of the two scenarios, local and remote, shows no significant differences in task performance or user rating except for the interaction time needed. The paired users were less than a second slower in the remote setup. User commentary indicates that the parquet flooring in the remote room was more slippery than the carpet in the main/local room which resulted in users being more careful in their movements. In addition, users dragged the cable behind them in the remote room, since the cable was not suspended from the ceiling. As a result, the speed difference might just be an environmental factor and not a factor of the user's location.

All pointing gestures performed well enough for users to consider the ability to talk to each other as a surplus. Qualitative ratings show that users feel equally co-present regardless of their actual, physical location. However, even the local setup did not achieve full ratings of co-presence from every user which could be explained by the reduced field of view of current VR headsets that limit the environmental awareness in VR compared to the real world or by the fact that the VR HMD and audio headset immerse the user so much that he feels as a part of the virtual world and tunes out reality. The participants show a slight preference for the remote setup since the do not need to worry about collisions. For applications that do not depend on direct contact between humans it might be therefore advantageous for users to collaborate from different locations.

6 Conclusion

VR allows users to collaborate in a virtual environment regardless of their physical location. The presented results of the user study show that an immersive VE enables collaboration regardless of the actual location of the different users. For the executed task almost no significant differences in a VR collaboration between two users in the same and separate physical locations could be found. User's performance, preferences and capability to collaborate seem to be equal in both setups.

The evaluation shows that a basic technical setup can already achieve this effect. The immersion into the virtual world and the feeling of co-presence might be increased even more with an improved and even more realistic avatar, haptics and environment. This opens up novel applications for collaboration in application domains such as education, design, diagnostics, and support. While these results show what is possible with regard to collaboration in VR, a key factor is the speed and latency of the network connection as it will severely influence the quality of the perceived co-presence. Further work is necessary to determine the minimum speed and maximum latency requirements to set a baseline for bringing collaboration in virtual reality to real-world applications. In addition, new techniques need to be developed that allow direct contact between humans in remote setups or collision avoidance mechanisms for local VR collaboration.

References

1. Arthur, K., et al.: Designing and building the PIT: a head-tracked stereo workspace for two users. In: 2nd International Immersive Projection Technology Workshop, pp. 11–12 (1998)
2. Beck, S., et al.: Immersive group-to-group telepresence. IEEE Trans. Vis. Comput. Graph. **19**(4), 616–625 (2013)
3. Benford, S., et al.: Collaborative virtual environments. Commun. ACM **44**(7), 79–85 (2001)
4. Bowman, D.A., et al.: An introduction to 3-D user interface design. Presence: Teleoper. Virtual Environ. **10**(1), 96–108 (2001). https://doi.org/10.1162/105474601750182342. ISSN 1054-7460
5. Cohen, J.: A power primer. Psychol. Bull. **112**(1), 155 (1992)
6. Gerhard, M., Moore, D.J., Hobbs, D.J.: Continuous presence in collaborative virtual environments: towards a hybrid avatar-agent model for user representation. In: de Antonio, A., Aylett, R., Ballin, D. (eds.) IVA 2001. LNCS (LNAI), vol. 2190, pp. 137–155. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44812-8_12

7. Kranstedt, A., et al.: Measuring and reconstructing pointing in visual contexts. In: Proceedings of the Brandial, pp. 82–89 (2006)
8. Mori, M., MacDorman, K.F., Kageki, N.: The uncanny valley [from the field]. IEEE Robot. Autom. Mag. **19**(2), 98–100 (2012)
9. Poupyrev, I., et al.: Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. In: Computer Graphics Forum, vol. 17, pp. 41–52. Wiley Online Library (1998)
10. Roth, D., et al.: Avatar realism and social interaction quality in virtual reality. In: 2016 IEEE Virtual Reality (VR), pp. 277–278. https://doi.org/10.1109/VR.2016.7504761
11. Salzmann, H., Jacobs, J., Froehlich, B.: Collaborative interaction in co-located two-user scenarios. In: Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments, JVRC 2009, pp. 85–92. Eurographics Association, Lyon (2009). ISBN 978-3-905674-20-0. https://doi.org/10.2312/EGVE/JVRC09/085-092
12. Steed, A., Schroeder, R.: Collaboration in immersive and non-immersive virtual environments. In: Lombard, M., Biocca, F., Freeman, J., IJsselsteijn, W., Schaevitz, R. (eds.) Immersed in Media, pp. 263–282. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-10190-3_11
13. Szalavári, Z., et al.: Studierstube: an environment for collaboration in augmented reality. Virtual Reality **3**(1), 37–48 (1998)
14. Wingrave, C.A., Bowman, D.A., Ramakrishnan, N.: Towards preferences in virtual environment interfaces. In: EGVE, vol. 2, pp. 63–72 (2002)

Bidirectional Communication for Effective Human-Agent Teaming

Amar R. Marathe(✉), Kristin E. Schaefer, Arthur W. Evans, and Jason S. Metcalfe

Human Research and Engineering Directorate,
United States Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
{amar.r.marathe.civ,kristin.e.schaefer-lay.civ,
arthur.w.evans20.civ,jason.s.metcalfe2.civ}@mail.mil

Abstract. The recent proliferation of artificial intelligence research is reaching a point where machines are able to learn and adapt to dynamically make decisions independently or in collaboration with human team members. With such technological advancements on the horizon, there will come a mandate to develop techniques to deploy effective human-agent teams. One key challenge to the development of effective teaming has been enabling a shared, dynamic understanding of mission space, and a basic knowledge about the states and intents other teammates. Bidirectional communication is an approach that fosters communication between human and intelligent agents to improve mutual understanding and enable effective task coordination. This session focuses on current research and scientific gaps in three areas necessary to advance the field of bidirectional communication between human and intelligent agent team members. First, intelligent agents must be capable of understanding the state and intent of the human team member. Second, human team members must be capable of understanding the capabilities and intent of the intelligent agent. Finally, in order for the entire system to work, systems must effectively integrate information from and coordinate behaviors across all team members. The combination of these three areas will enable future human-agent teams to develop a shared understanding of the environment as well as a mutual understanding of each other, thereby enabling truly collaborative human-agent teams.

Keywords: Automation · Autonomy · Robot · Mixed-initiative teams
Human automation interaction · Bidirectional communication

1 Introduction

Integration of advancing intelligent technologies on the battlefield will change the very nature of the tasks Warfighters need to perform. This, in turn, will require the evolution of different skill sets and capabilities, which will thus impact the precise needs for those Warfighters. A research and development approach is needed that conceives of not only the potential capabilities of these future intelligent technologies, but the potential for completely novel interactions among heterogeneous teams of Warfighters and intelligent agents, and reconceives approaches and requirements for training.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 338–350, 2018.
https://doi.org/10.1007/978-3-319-91581-4_25

Within these teams, the human is integral to decision-making, including adapting requirements to dynamic events, and completion of the overall mission. As such, there is a need to have the human-agent team perform as well as human-only teams but with the potential to provide additional support for advanced mission directives. The benefits of integrating intelligent agents into a human team include the potential for greater team resilience with robust, adaptive performance; faster, dynamic human-agent teaming (HAT) reconfiguration to match capabilities to mission requirements; faster, more informed team decision making; and reduced numbers and risk to human team members.

Developing advanced intelligent technologies that are capable of functioning as a natural teammate is a critical challenge for the research community. Research into successful human teams has shown that performance outcomes are not simply a sum or an average of the performance of the individuals. Instead, emergent properties are the result of the interaction of the components of the team, which cannot be reduced to or described wholly in terms of the individual elements of the system considered in isolation. Team performance often breaks down because of problems with emergent team states and process such as insufficient communications, misunderstanding of team goals, undefined team responsibilities or lack of shared mental models, and conflict [1, 2].

Moving from human-only teams, to mixed-agent teams composed of humans, intelligent software agents, embodied agents (e.g., robots), and networked sensors adds complexity that may not be completely comprehended today; this especially in uncertain informational environments with limited or unreliable communications. To develop effective mixed agent teams, humans and agents must be allowed to work in disparate dimensions (time, space, world views, representations, mental models, etc.), but also capable of seamlessly synchronizing for collective action. For example, intelligent agents will process information, reason, and make decisions at scales beyond that of humans in both time and scope, and yet we will want to include humans in the decision-making loop for many if not most battlefield decisions on account of their superior abilities to adapt and develop abstract understandings in the face of novelty and uncertainty. This brings to light the open research question of how to capitalize on the individual advantages of both humans and agents, and simultaneously enhance the performance of the collective group.

While this is largely an open question, part of the solution to these issues is to developing methodologies to enable bidirectional communication between the human and intelligent agents. Successful human teams often can communicate effectively, have team members who possess knowledge of each other's skills and capabilities that allow them to anticipate each other's actions, and can interpret and assess the environmental constraints on the task at hand. Building a robust capability for resilient bidirectional communication is believed to be an important approach to developing these same properties in human-agent teams by facilitating increased mutual understanding between human and intelligent agents. In this session, we discuss three broad challenge areas that must be addressed to enable the development highly functioning human-agent teams.

Agents Understanding Humans. Intelligent agents must be provided with information necessary to develop an understanding of their human team members. This requires leveraging wearable (and non-wearable) technologies to develop human assessment tools capable of generating high-resolution, real-time, predictions of an individual's internal and external behavioral and performance dynamics across a variety of environments. The specific types of predictions necessary will depend on how that information is to be used, the current task, and the current environment. Two uses of the information are considered in the sections that follow.

Humans Understanding Agents. Humans must be capable of understanding the capabilities and intent of their intelligent agent teammates. This requires developing adaptive interfaces for the intelligent agents that provide the human with the appropriate level of detail regarding the intelligent agent decisions or behavior. These adaptive interfaces will leverage human assessment tools described in the previous section to customize the displays for the current state of a particular human. These displays must convey information about the intelligent agent behavior or decisions with consideration for issues such as transparency, trust, and team situation awareness.

Joint Human-Agent Teamwork. Integration of intelligent agents and Warfighters requires the capability to deliver appropriate information to both the human and intelligent agent at the appropriate time. This necessitates the need for developing integration principles and approaches that dynamically accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved performance. This type of integration requires insight into the current and future states of each individual agent, which, for humans, should come from the human assessment tools described in the Sect. 2.

2 Agents Understanding Humans

Human-agent team performance is not limited by computing power, but by the ability for computer or embodied intelligent agents to understand humans. This is evidenced by overwhelming majority of current systems that assume fixed, stereotyped human input that is geared toward an "average" human. These systems assume that the quality of human input, which is often presumed to be of low noise and high accuracy, will be static over time and across individuals. We consider this to be a fundamentally flawed assumption because of inter and intra-individual differences in humans and agents. To account for human variability during HAT, continuous, real-time human assessment technologies, which will combine human and environmental sensing technologies with advanced analytics, will provide the foundational elements for future systems to generate high-resolution, moment-to-moment, predictions of individual's internal and external behavioral and performance dynamics across training and operational environments. Fundamentally, this capability will enable future systems to move from an approach of *mitigating* the effects of human variability to one that *embraces and predictively capitalizes* on that variability.

The US Department of Defense, Defense Science Board [3] report on autonomy suggests that the future value of unmanned systems lies not in the direct replacement of any one particular human operator, but rather in their contribution to overall human-system collaboration (e.g., the capability to extend and complement human capability without degradation due to factors such as fatigue, stress, lack of attention, situation awareness, amongst others). Further, calibrating inter-agent trust is considered essential to this collaboration. Some identified critical gaps to developing trust in HAT include the impact of human states (stress, fatigue, and attentional control), cognitive factors (understanding of technology, ability to use or interact with technology and expectance), and emotive factors (confidence, attitudes, satisfaction and comfort) on teaming with respect to task-specific environmental factors including risk, uncertainty, task type, context, and the physical environment [4]. While a number of current research efforts are underway to better understand and quantify these critical gaps, continued research will advance our understanding of the human performance during real-time, real-world operations, as well as advance adaptive autonomy technologies leading to more advanced, collaborative teaming.

Near and mid-term research will focus on leveraging sensing technologies to enable high fidelity, omnipresent prediction of behaviors and intentions that can account for continuous changes in an individual's physical, cognitive, and social states (examples from these three types of states include stress, workload, fatigue, task difficulty, trust, and situation awareness). The goal is to enable the exploitation of the array of sensors and information streams to predict human performance dynamics with sufficient resolution and robustness to adapt systems in manners to directly enhance performance.

The current state of the field has demonstrated unparalleled advancements in sensor and analysis technologies that provide new insights into different facets of human psychology, physiology, behavior, and performance. For example, advances in neuroscientific tools have revealed novel discoveries on how differences in brain function influence precise human behaviors [5, 6]. Advances in social and environmental sensing tools and have provided unprecedented insights into patterns of gross human social behaviors [7], while advances in biochemical or fluid sensing (i.e., blood, sweat, and tears) are providing unique insights into the continuous dynamics of internal human states and traits. More generally, advances in wearable devices have enabled the tracking of a wide range of factors including activity, sleep patterns, and various physiological parameters (see [8] for an overview).

Even with this broad explosion in sensing technologies, there is a lack of understanding regarding the factors that influence variability in human performance, as well as an inability to develop predictive algorithms that account for this variability. This has prevented a similar explosion in human assessment technologies capable of providing robust predictions regarding health and performance. As we develop a better understanding of human variability, a combination of sensing, analytic, and enabling technologies must be leveraged to assess and predict human states, behaviors, intentions, and performance. This capability will provide the foundation for a broad range of individualized, adaptive technologies across a variety of domains. Understanding the human team member in such detail can then be used to influence autonomy design. A sampling of critical technologies include:

- Critical Sensing Technologies, including wearable and non-wearable sensors, provide novel insights into the human and their local environment that could be communicated to an intelligent agent. These types of sensors provide data regarding intended behaviors, unintended behaviors, physiology, brain activity, subjective experiences, and social interactions; as well as task constraints and a wide range of environmental and societal factors. These sensing technologies arise from fields including quantified self; brain-computer interface; human augmentation; bioacoustic sensing; electrovibration; speech recognition and translation; gamification; wearable user interfaces; mobile health monitoring; gesture control; activity streams; and pervasive computing. The sensors fall in the following general categories of on-body sensing (e.g., watches, shirts, cell phones, glasses), off-body sensing (e.g., cameras, computers, fixed-placement), and their combinations for use in network-based sensing (e.g., network interactions, team and organizational performance, societal events).

- Critical Analytic Technologies merge and interpret data from a wide variety of human, task, and environmental sources to communicate higher-level goals and understanding of the mission space and human team members. These arise from fields including: brain-computer interface, affective computing, prescriptive analytics, natural-language response, big data, complex-event processing, content analytics, location intelligence, social network dynamics, and predictive analytics.
- Critical Enabling Technologies provide an integration framework for consolidating these data into a tractable repository for analysis by augmenting current network and computational technologies. Examples of these technologies will rely upon wireless local area networking, wireless bridge technologies, frequency domain equalization optimization, cloud computing, and smart antennae. These types of technologies provide fundamental solutions for collecting, storing and analyzing sensed data in real-time, thereby enabling predictions of human state and performance to be conveyed to intelligent agents.

3 Humans Understanding Agents

Intelligent technologies, such as embodied agents, are not yet operating as teammates in the field. They are by and large either teleoperated or supervised tools that possess insufficient shared understanding to independently adapt to the benefit of the team [9]. However research efforts are underway to advance intelligence architectures for independent and collaborative decision-making capabilities that can account for uncertain and dynamic environments [10, 11]. As agents become more sophisticated and independent, it is critical for their human counterparts to understand their behaviors, the reasoning process behind those behaviors, and the expected outcomes to properly calibrate their trust in the systems and make appropriate decisions [12, 13]. This is essential to the team effort because people will question the accuracy and effectiveness of agents' actions if they have difficulty understanding the state or status of the agent [14–16] and the reasoning behind specific actions or behaviors [17–19]. As such, if user expectations do not match agent actions (even if the actions of the agent are optimal and appropriate decisions), then trust can degrade to the degree in which the user will misuse or disuse the agent [20, 21].

These limitations with how humans understand agents can be substantial impediments to overall system, task, and team performance. To overcome these limitations, future systems will leverage detailed assessments of the human, as described in Sect. 2, to dynamically adapt and tailor the interface to fit the informational needs dictated by the precise teaming function or functions to be shared between the autonomous agent and human. These adaptive interfaces will ensure that the human is given the information needed at the appropriate time and in the appropriate scale and frame of reference to maximize overall performance.

Recent research efforts examine the integration of spatial and temporal context into artificial intelligence (AI) development. Semantic mapping is used to label objects in the world in order to assign high-level information to decisions, such as those needed for communicating decision related to path generation and mapping [22, 23]. Computer vision [24, 25] and natural language [26] have been used to develop probabilistic models that can provide an abstraction of the environment and better support intelligent agent-to-human communication needs. Research into temporal context and AI has looked at time perception in decision-making [27]. This work is important because it helps in the development of shared situation awareness by enabling the capability of interpreting the state or recognizing the current situation by observing the partial or entire state history at any time.

This link between AI and the human's understanding of the agent can directly influence bidirectional communication needed for appropriate trust development through what Lee and See [12] termed the 3 P's: purpose, process, and performance. Purpose deals with the degree to which the agent-driven automation is being used according to the designer's intent. Process deals with the question of whether the algorithm of the automated system is appropriate for a given task. Performance deals with system reliability, predictability, and capability. Lee [28] proposed that to increase system transparency to the human team member, the system's 3Ps, as well as the history of the 3Ps, should be visible to the operator. However, the presentation should be in a simplified form (e.g., integrated graphical displays) so the operator is not overwhelmed by the additional information he/she needs to process [29–31]. From this basis, Chen et al. [18] developed the Situation awareness-based Agent Transparency (SAT) model, which identifies and organizes the information that an agent needs to share with the human teammate to support their situation awareness, trust, and appropriate reliance upon the agent. This effort demonstrates the potential for developing interfaces that dynamically adapt to the state and needs of the team members (possibly quantified by wearable technologies), the task at hand, and the situational context to provide necessary and sufficient explanations to enhance team performance while simultaneously building and maintaining team cohesion.

As an extension to the transparency issue, trust in automation (TiA) has long been considered central in influencing the way a human user interacts with an automation; if TiA is too high there will be overuse, if TiA is too low there will be disuse [12]. TiA is an important construct that undoubtedly affects human user interaction behaviors. However the relationship between TiA and human behavior is complex and not currently fully understood. Further, relevant to immediate real-world applications for improving team performance, TiA measurement has most commonly leveraged subjective metrics [32] occurring after an interaction, with only some more recent research efforts looking at behavioral indicators and physiological markers of trust to map to potential real-time implications [13, 21, 33–35]. However, current research has suggested that transparent communication of agent intent can support collaboration and in turn calibrate trust and reliance on the system [36]. The specific criteria for this communication are still being developed.

Further, from an engineering standpoint, it is not yet possible to define an objective function based on TiA that would adequately define how control authority should shift dynamically between operator and automation. By contrast, specific behaviors are readily observed and measured in real time and do not have the confounding effect of inferring psychological causality. Recent studies have shown that an operator's decision to toggle control between human and autonomy drivers may be predicted on the basis of behavior, physiological, and environmental factors [37]. These predictions can integrated with dynamic estimates of performance to calibrate trust-based decisions [35].

As the future of human-agent interaction moves towards interdependent teaming initiatives, developing efficient complex decision-making processes is an essential part of the design and development of autonomous agents. The process by which agents make decisions is still an open research question when operating in uncertain environments with unknown data. Current research has demonstrated the importance and relevance to why understanding agent decision-making processes is relevant to performance. It has been shown that humans and intelligent-agents faced with the same circumstances will not make the same decisions, even under the same set of apparent constraints, nor will they necessarily have the same consequences resulting from those decisions. Perelman and colleagues [38] found that when comparing human and algorithmic path planning, there is more than one 'human way' of solving a planning problem which may or may not match an algorithm. The potential mismatch of solutions, without explanation, could result in significant degradation of team process. Human teammates may reduce their interaction with or outright ignore intelligent agents regardless of how correct the solutions are which they provide.

4 Joint Human-Agent Teamwork

In order to develop effective mixed-agent teams, technologies for inferring motivations, predicting behavior and reasoning about the environment must be incorporated into a closed-loop system that can initiate individualized interventions to improve team performance. In general, humans are able to adapt to the complexities and dynamics of real-world operational environments to a degree unmatched by current forms of autonomy. However, humans simply cannot process the full amount of information available or understand the reasoning of complex, intelligent agents. Thus, we must develop novel integration principles and approaches that leverage advanced sensing, data processing, and dynamic inferential tools in a way that can enable us to accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved decision making.

It has long been understood that, though automation can execute predictable, well-defined procedures with superior speed and reliability, humans are far superior at tasks that require inductive reasoning and adaptation to novel or changing information [39–

41]. As a result, system integrators have developed a wide range of approaches to supplement intelligent agent autonomy with human inputs to increase resilient and robust performance within complex, dynamic, and uncertain environments. Yet most systems-level, human-centric design approaches have treated the human as most appropriately positioned at the peak of the command hierarchy (c.f. [42–45]) rather than as a fully collaborative teammate [46]. That is, while these approaches have not always required the user to give continuous control or decision inputs and corrective feedback, when the human input has been available, it has usually been integrated as the *de facto* correct solution, which is not always true. Treating the human as the ultimate and final authority is a premise based on either an explicit mandate [42] or on an assumption that human influence would guarantee optimal behavior in situations where an automated agent is uncertain or otherwise compromised.

An ethical debate has emerged as to whether or not a person should retain overarching decision-making authority and perhaps more importantly, accountability [47]. Are there times that an intelligent-agent should make the decision rather than the user? We do not see the answer to this question as black-and-white. Variables such as context and limitations associated with the intelligent-agent's ability to physically detect and make sense of the environment, as well as to infer the intent of the human teammate, will critically impact whether a response is appropriate or inappropriate. Human intervention may be appropriate in one condition and inappropriate in another. We acknowledge the complex and sometimes emotional aspects of this debate. This transfer process is difficult to manage and requires balancing control allocation within the team structure while maintaining team commitments and supporting large-scale interactions with multiple agents. Reengagement for a person can sometimes be difficult and may impact safety and mission effectiveness due to a user's situation awareness, workload, and abilities. We join those who have argued that adherence to this premise has limited how well human inputs have been integrated with autonomous systems [46, 48, 49].

There has been considerable research into mitigating the potential impact of human variability and performance failures on HAI systems. Unfortunately, the majority of these approaches have only succeeded in limited and controlled contexts, and have not been widely adopted for real-world use. We argue that this is due, at least in part, to adherence to the axiomatic premise that the human should be the ultimate and final authority; with failure to fully account for the dynamic strengths and vulnerabilities of the human team member being a critical design outcome of this belief.

Recent efforts have proposed the Privileged Sensing Framework (PSF), an evolved approach that treats the human as a special class of sensor rather than as the absolute command arbiter [50]. This approach is based on the concept of appropriately 'privileging' information during the process of integrating information from human and autonomy team members by bestowing advantages, special rights, or immunities based on the characteristics of each individual agent (on the basis of data from wearable and non-wearable sensors), the task context, and/or the performance goals. Indeed, treating the human as a privileged sensor deviates from the established central axiom of human-centered automation [51]. However, we view this departure as an important evolutionary step beyond substitution-based function allocation methods [52] and in alignment with notions of human-automation interactions that capture a more authentic essence of natural teaming behavior [46, 53, 54].

5 Session Details

Focusing efforts on developing bidirectional communication as a critical capability may be an essential approach to human-agent teaming that seeks to develop common ground and shared understanding between human and intelligent agent team members. In this session, we discuss current research and gaps within three focus areas. First, intelligent agents must be capable of understanding the state and intent of the human team member. This area of research is largely focused on sensing the human and providing the intelligent agents with information related to estimated state and expected performance of that human team member. Second, human team members must be capable of understanding the intelligent agent. This area of research is focused on conveying information about the intelligent agent back to the human team member in a manner most appropriate to that human team member at a given point in time. This means, that many of these approaches will involve adapting the information conveyed to the human based on the information the agent has about that team members current state. Finally, in order for the entire system to work, we must develop closed loop solutions to effectively integrate information from and coordinate behaviors across human and intelligent agent team members. This area of research focuses on developing integration principles and approaches that accentuate the strengths of individual humans and agents while mitigating relative weaknesses for improved team performance. Within this session, we have three talks that cover a subset of these areas.

5.1 A Maximum Likelihood Method for Estimating Performance

Jonroy Canady and colleagues provide an example of estimating human decisions that may be used to improve human agent teaming within the context of a joint human-agent target identification paradigm [55]. In this work, they present a method for creating unbiased and accurate estimates of human target-detection performance that could be used to better integrate the human decisions with those of the autonomous agents. This study provides an example of how information sensed from human responses can be conveyed to an intelligent agent. Within the bidirectional communication framework, an intelligent agent teammate would then leverage that information to adapt their behavior in a manner that improves overall system performance.

5.2 Quantifying Human Decision-Making

Kristin Schaefer and colleagues address the importance of human spatial decision-making as it applies to the development of appropriate human-agent team communication [56]. In this work, they aim to quantifying human spatial decision-making because if human behavior does not match the robots' models or expectations, there can be a degradation in trust that can impede team performance. Degradation in trust may only be mitigated through explicit communication which are needed to develop common ground and a shared understanding. To reach this end, this study identifies divergence in planning and action times and behaviors as well as detects differences in local versus global decision-making processes needed to predict complex decisions within the confines of increasing environmental complexity and amount of prior knowledge about the task. It supports agents understanding humans in that findings from quantifying human spatial decision-making can advance the technical capabilities of a robot to more accurately perceive and interpret human team member behavior. The next step in the research will support humans understanding agents as these findings are compared to robot solutions. Where disparities exist between the resultant robot and human behaviors, bidirectional communication can be used as a means to achieve an optimal solution collaboratively.

5.3 The Role of Psychophysiological Measures Within Mixed-Initiative Teams

Kim Drnec and colleagues extend their previous work on characterizing behavioral, physiological and task-based factors that influence trust-in-autonomy by examining the efficacy of a real-time system that uses a combination of behavioral and psychophysiological data from a human driver to foster more appropriate use of autonomous driver assist technologies [57]. This study provides an example of a system using bidirectional communication between human and intelligent agent to improve overall performance. The system uses information from wearable sensors about behavior, and physiology and combines that with environmental and task based factors to predict human intent to toggle control (*Agent Understanding Warfighter*). The system then uses that information about the human driver to provide customized feedback to assist the driver in making an appropriate decision regarding whether or not to toggle driving control (*Warfighter Understanding Agent*). The overall system also shows a simple case of using bidirectional communication to share information between a human and intelligent agent to maximize performance.

Acknowledgements. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Donaldson, M.S., Corrigan, J.M., Kohn, L.T., et al.: To Err is Human: Building A Safer Health System. National Academies Press, Washington D.C. (2000)
2. Salas, E., Stagl, K.C., Burke, C.S., Goodwin, G.F.: Fostering team effectiveness in organizations: toward an integrative theoretical framework. In: Nebraska Symposium on Motivation, p. 185 (2007)
3. Murphy, R., Shields, J.: The role of autonomy in DoD systems. Def. Sci. Board (2012)

4. Schaefer, K.E., Chen, J.Y., Szalma, J.L., Hancock, P.A.: A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum. Factors **58**, 377–400 (2016)
5. Detection of functional brain network reconfiguration during task-driven cognitive states. Neuroimage. **142**, 198–210 (2016)
6. Garcia, J.O., Brooks, J., Kerick, S., Johnson, T., Mullen, T.R., Vettel, J.M.: Estimating direction in brain-behavior interactions: Proactive and reactive brain states in driving. NeuroImage **150**, 239–249 (2017)
7. Kalia, A.K., Buchler, N., DeCostanza, A., Singh, M.P.: Computing Team Process Measures From the Structure and Content of Broadcast Collaborative Communications. IEEE Trans. Comput. Soc. Syst. **4**, 26–39 (2017)
8. Bonato, P.: Wearable sensors and systems. IEEE Eng. Med. Biol. Mag. **29**, 25–36 (2010)
9. Chen, J.Y., Haas, E.C., Barnes, M.J.: Human performance issues and user interface design for teleoperated robots. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **37**, 1231–1245 (2007)
10. Dean, R.M.S.: Common world model for unmanned systems. In: Unmanned Systems Technology XV, p. 87410O. International Society for Optics and Photonics (2013)
11. Dean, R.M.S., Oh, J., Vinokurov, J.: Common world model for unmanned systems: phase 2. In: Unmanned Systems Technology XVI, p. 90840I. International Society for Optics and Photonics (2014)
12. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors J. Hum. Factors Ergon. Soc. **46**, 50–80 (2004)
13. de Visser, E.J., Cohen, M., Freedy, A., Parasuraman, R.: A design methodology for trust cue calibration in cognitive agents. In: Shumaker, R., Lackey, S. (eds.) VAMR 2014. LNCS, vol. 8525, pp. 251–262. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07458-0_24
14. Bitan, Y., Meyer, J.: Self-initiated and respondent actions in a simulated control task. Ergonomics **50**, 763–788 (2007)
15. Seppelt, B.D., Lee, J.D.: Making adaptive cruise control (ACC) limits visible. Int. J. Hum.-Comput. Stud. **65**, 192–205 (2007)
16. Stanton, N.A., Walker, G.H., Young, M.S., Kazi, T., Salmon, P.M.: Changing drivers' minds: the evaluation of an advanced driver coaching system. Ergonomics **50**, 1209–1234 (2007)
17. Linegang, M.P., Stoner, H.A., Patterson, M.J., Seppelt, B.D., Hoffman, J.D., Crittendon, Z.B., Lee, J.D.: Human-automation collaboration in dynamic mission planning: a challenge requiring an ecological approach. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 2482–2486. SAGE Publications, Los Angeles (2006)
18. Chen, J.Y., Procci, K., Boyce, M., Wright, J., Garcia, A., Barnes, M.: Situation awareness-based agent transparency. Army Research Laboratory (2014)
19. Chen, J.: Advances in Human Factors in Robots and Unmanned Systems (2018)
20. Schaefer, K.: The perception and measurement of human-robot trust (2013)
21. Schaefer, K.E., Straub, E.R.: Will passengers trust driverless vehicles? Removing the steering wheel and pedals. In: 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), pp. 159–165. IEEE (2016)
22. Kostavelis, I., Gasteratos, A.: Semantic mapping for mobile robotics tasks: a survey. Robot. Auton. Syst. **66**, 86–103 (2015)
23. Wolf, D.F., Sukhatme, G.S.: Semantic mapping using mobile robots. IEEE Trans. Robot. **24**, 245–258 (2008)

24. Heitz, G., Koller, D.: Learning spatial context: using stuff to find things. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 30–43. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88682-2_4
25. Meger, D., Forssén, P.-E., Lai, K., Helmer, S., McCann, S., Southey, T., Baumann, M., Little, J.J., Lowe, D.G.: Curious george: an attentive semantic robot. Robot. Auton. Syst. **56**, 503–511 (2008)
26. Walter, M.R., Hemachandra, S., Homberg, B., Tellex, S., Teller, S.: Learning semantic maps from natural language descriptions (2013)
27. Maniadakis, M., Trahanias, P.: Time models and cognitive processes: a review. Front. Neurorobot. **8**, (2014)
28. Lee, J.: Trust, trustworthiness, and trustability. In: Presentation at the Workshop on Human Machine Trust for Robust Autonomous Systems (2012)
29. Cook, M.B., Smallman, H.S.: Human factors of the confirmation bias in intelligence analysis: decision support from graphical evidence landscapes. Hum. Factors **50**, 745–754 (2008)
30. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity: incorporating system reliability information into an automated combat identification system. Hum. Factors **53**, 338–355 (2011)
31. Tufte, E.R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire (2001)
32. Schaefer, K.E.: Measuring trust in human robot interactions: development of the "*Trust Perception Scale-HRI*". In: Mittu, R., Sofge, D., Wagner, A., Lawless, W.F. (eds.) Robust Intelligence and Trust in Autonomous Systems, pp. 191–218. Springer, Boston, MA (2016). https://doi.org/10.1007/978-1-4899-7668-0_10
33. Drnec, K., Metcalfe, J.S.: Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In: Schmorrow, D.D.D., Fidopiastis, C.M.M. (eds.) AC 2016. LNCS (LNAI), vol. 9744, pp. 157–167. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39952-2_16
34. Gremillion, G.M., Metcalfe, J.S., Marathe, A.R., Paul, V.J., Christensen, J., Drnec, K., Haynes, B., Atwater, C.: Analysis of trust in autonomy for convoy operations. In: SPIE Defense + Security, p. 98361Z. International Society for Optics and Photonics (2016)
35. Nothwang, W.D., Gremillion, G.M., Donavanik, D., Haynes, B., Atwater, C., Canady, J., Metcalfe, J.S., Marathe, A.R.: Multi-sensor fusion architecture for human-autonomy teaming. In: 2016 9th International Symposium on Resilient Control Systems (ISRCS). IEEE (2016)
36. Schaefer, K.E., Straub, E.R., Chen, J.Y., Putney, J., Evans, A.: Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cogn. Syst. Res. **46**, 26–39 (2017)
37. Metcalfe, J., Marathe, A., Haynes, B., Paul, V., Gremillion, G., Drnec, K., Atwater, C., Estepp, J., Lukos, J., Carter, E., et al.: Building a framework to manage trust in automation. In: Micro- and Nanotechnology Sensors, Systems, and Applications IX, p. 101941U. International Society for Optics and Photonics (2017)
38. Perelman, B., Evans III, A., Schaefer, K.: Mental model consensus and shifts during navigation system-assisted route planning. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 1183–1187. SAGE Publications, Los Angeles (2017)
39. Fitts, P.M.: Human engineering for an effective air-navigation and traffic-control system (1951)
40. Sheridan, T.B.: Function allocation: algorithm, alchemy or apostasy? Int. J. Hum.-Comput. Stud. **52**, 203–216 (2000)
41. Cummings, M.M.: Man versus machine or man + machine? IEEE Intell. Syst. **29**, 62–69 (2014)

42. Billings, C.E.: Human-centered aircraft automation: a concept and guidelines (1991)
43. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)
44. Fong, T., Thorpe, C., Baur, C.: Multi-robot remote driving with collaborative control. IEEE Trans. Ind. Electron. **50**, 699–704 (2003)
45. Abbink, D.A., Mulder, M., Boer, E.R.: Haptic shared control: smoothly shifting control authority? Cogn. Technol. Work **14**, 19–28 (2012)
46. Woods, D.D., Branlat, M.: Hollnagel's test: being "in control" of highly interdependent multi-layered networked systems. Cogn. Technol. Work **12**, 95–101 (2010)
47. Schaefer, K.E., Brewer, R.W., Putney, J., Mottern, E., Barghout, J., Straub, E.R.: Relinquishing manual control: collaboration requires the capability to understand robot intent. In: 2016 International Conference on Collaboration Technologies and Systems (CTS), pp. 359–366. IEEE (2016)
48. Woods, D.D.: Cognitive technologies: the design of joint human-machine cognitive systems. AI Mag. **6**, 86 (1985)
49. Cummings, M., Clare, A.: Holistic modelling for human-autonomous system interaction. Theor. Issues Ergon. Sci. **16**, 214–231 (2015)
50. Marathe, A.R., Metcalfe, J.S., Lance, B.J., Lukos, J.R., Jangraw, D., Lai, K.T., Touryan, J., Stump, E., Sadler, B.M., Nothwang, W., McDowell, K.: The privileged sensing framework: A principled approach to improved human-autonomy integration. Theor. Issues Ergon. Sci. **19**(3), 283–320 (2018)
51. Billings, C.E.: Aviation automation: The search for a human-centered approach. Lawrence Erlbaum Associates Inc., Mahwah (1997)
52. Dekker, S.W., Woods, D.D.: MABA-MABA or abracadabra? Progress on human–automation co-ordination. Cogn. Technol. Work **4**, 240–244 (2002)
53. Lyons, J.B.: Being Transparent about Transparency (2013)
54. Chen, J.Y., Barnes, M.J.: Human–agent teaming for multirobot control: a review of human factors issues. IEEE Trans. Hum.-Mach. Syst. **44**, 13–29 (2014)
55. Canady, J.D., Marathe, A.R., Herman, D.H., Files, B.T.: A maximum likelihood method for estimating performance in a rapid serial visual presentation target-detection task. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 383–392. Springer, Heidelberg (2018)
56. Schaefer, K.E., Perelman, B.S., Brewer, R.W., Wright, J., Roy, N., Aksaray, D.: Quantifying human decision-making: implications for bidirectional communication in human-robot teams. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 361–379. Springer, Heidelberg (2018)
57. Drnec, K., Gremillion, G.M., Donavanik, D., Canady, J.D., Atwater, C.S., Haynes, B.A., Marathe, A.R., Carter, E.C., Metcalfe, J.S.: The role of psychophysiological measures as implicit communication within mixed-initiative teams. In: Schmorrow, D., Fidopiastis, C.M. (eds.) VAMR 2018. LNCS 10909, pp. 299–313. Springer, Heidelberg (2018)

PaolaChat: A Virtual Agent with Naturalistic Breathing

David Novick(✉), Mahdokht Afravi(✉), and Adriana Camacho(✉)

The University of Texas at El Paso, El Paso, TX 79968, USA
novick@utep.edu, mmafravi@gmail.com, caro4874@gmail.com

Abstract. For embodied conversational agents (ECAs) the relationship between gesture and rapport is an open question. To enable us to learn whether adding breathing behaviors to an agent similar to SimSensei would lead users interacting to perceive the agent as more natural, we built an application, called Paola Chat, in which the ECA could display naturalistic breathing animations. Our study had two phases. In the first phase, we determined the most natural amplitude for the agent's breathing. In the second phase, we assessed the effect of breathing on the users' perceptions of rapport and naturalness. The study had a within-subjects design, with breathing/not-breathing as the independent variable. Despite our expectation that increased naturalness from breathing would lead users to report greater rapport in the breathing condition than in the not-breathing condition, the study's results suggest that the animation of breathing appears to neither increase nor decrease these perceptions.

Keywords: Embodied conversational agents · Human-agent dialog
Dialog system

1 Introduction

The relationship between embodied conversational agents' (ECAs) [1] gestures (see e.g., [2]) and rapport (see e.g., [3, 4]) is a currently active research question. While some studies have reported the effect of users' perception of extraversion on gesture amplitude [5], other studies reported that gesture amplitude may not affect users' perception of rapport [6]. This disparity suggests that the naturalness of ECAs' gestures may be a significant factor in shaping users' perceptions. Indeed, not all ECAs have the same level of naturalness of behavior.

The building of human-ECA rapport is increasingly important as ECAs take on more meaningful roles, including serving as a means of diagnosing PTSD [7]. SimSensei, the PTSD-diagnosis agent, while having excellent animation of facial features, appeared to have a static torso, which might give the impression that she is holding her breath or not breathing between her utterances. Therefore, we sought to answer the question of whether adding naturalness, in this case for breathing, would lead to higher perceptions of rapport.

To this end, we built an application that would enable us to learn whether adding breathing behaviors to a similar agent would lead humans to perceive the agent as more

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 351–360, 2018.
https://doi.org/10.1007/978-3-319-91581-4_26

natural and to develop a higher level of rapport with the agent. The application we developed, called Paola Chat, featured an ECA named Paola that resembled SimSensei but was able to display naturalistic breathing animations. The animations were based on a simple model of respiration and an empirical study of the perceived naturalism of breathing amplitude. The application enabled study of whether users' perceptions of the ECA's naturalness would increase based on the varying frequency and amplitude of the ECA's breathing during the conversation.

2 Implementation

This study comprised two phases: a preliminary phase in which we sought to avoid the possible effects non-standard amplitude could have on perceptions of naturalness (cf., [6]) and a second phase in which we assessed the effect of breathing on users' perceptions of rapport.

2.1 Phase 1: Naturalistic Amplitude

In the first phase, we sought to find the most natural amplitude of gestures for perceptions of naturalness (cf., [6].) To provide the Paola Chat agent with amplitude of breathing that would be as natural as possible, we conducted an empirical evaluation of human perceptions of the agent with different amplitudes for the breathing animation. We prepared five brief animations of the ECA's breathing, ranging from static (not breathing) to exaggerated breathing. We sought to have the agent show breathing with amplitude large enough to be salient but not so large as to appear unnatural or distracting. Each participant viewed the five representative animations as an introduction to the task. Each participant then saw and rated 30 animations for naturalness, presented in random order, using a five-point Likert scale.

Unsurprisingly, we found that that the animation rated the most natural was the medium-amplitude animation, which had a 50% amplitude from the extreme. Figure 1 compares Paola's breathing at the last point of inhale (just before the transition to exhale) with Paola's breathing at the last point of exhale (just before the transition to inhale) in this medium amplitude; the differences apparent in this static figure are subtle.

2.2 Phase 2: Effect on Perception of Rapport

With the amplitude determined, in the study's second phase we assessed the effect of breathing on the users' perceptions of rapport. To this end, we implemented Paola Chat, which is a fully automated conversational agent rather than a Wizard-of-Oz system. We designed Paola to resemble SimSensei as much as possible. Paola, displayed as a life-sized person projected on the wall in UTEP's immersion laboratory, was seated in a large chair, with her legs crossed and her hands resting on her lap, resting on the chair's armrests, or making gestures while speaking (see Fig. 2). Users' interactions with Paola consisted of two back-to-back conversations on the topics of vacations and movies. For example, in the movie conversation, Paola asked "Have you seen any movies lately?"

Fig. 1. Paola's inhale before transition (left) to exhale (right).

Paola would then interpret the user's response using keyword recognition (i.e., if the user answered that he or she had, Paola would ask for details about the movie, but if the user had responded in the negative, Paola would segue to another question about the user's favorite movie).

Fig. 2. A person interacts with the Paola Chat agent.

Developing this application required accepting a wide range of dialogue input and generating relevant responses. Paola Chat was developed with the UTEP AGENT system [8], which is capable of accepting a wide range of utterances, called wildcards, irrespective of topic or word choice. A major difficulty was that the study needed to have Paola's utterances generate extended responses from the users so that they could observe

Paola both while she was talking and while she was listening. But, consequently, the system must be able to generate responses that are generic enough to keep the conversation from seeming one-sided or disconnected. Some questions were posed as “yes” or “no” response questions, in which case the dialogue tree would converge back to a certain point to naturalize the flow of utterances generated by Paola.

Breathing Model. The agent needed a function to control the breathing with respect to three constraints. First, the agent should not be speaking while displaying an inhaling animation. Second, the agent should appear to take breaths of natural length between utterances. Third, the agent’s frequency (in addition to amplitude) of breathing should be perceived as natural.

Breathing includes the states of inhaling or exhaling, and their transitions [8]. Our system required the development of a model that could represent these states smoothly, as well as having amplitude, oscillation, and frequency. In our model, the breathing state depends on the amplitude and oscillation. The amplitude represents the y-value on a graph and visually represents how much the ECA’s torso would expand. The oscillation represents the x-value on a graph in radians. The oscillating state (i.e., the wave) moves depending on the frame rate of the animation and the frequency per frame. The frequency was set to a fixed value per frame, adjusted for frame drops because Unity sometimes skips frames. During the interaction, the oscillations varied between 100 and 0, as an effect of the sine wave function:

$$\text{Breathing State} = \text{Amplitude} + (\text{Amplitude} \times \sin(\pi \times \text{Frequency}))$$

The breathing oscillating function ran in a cycle in which the agent would either inhale (breathing state = 0) or exhale (breathing state = 100). The cycle was interrupted only when the agent was about to speak, to portray an in-breath before speech. Figure 3 shows the breathing oscillation function: the x-axis is the oscillation of sin(πx), where x is the frequency, and the y-axis is the changes in the wave function for the values of the updating breathing state, with 0 the lowest point of exhale and 100 as the highest point of inhale. Figure 4 depicts the transitions between the states in the breathing oscillating function.

Application Dialog. We deployed Paola Chat’s breathing model in a pair of conversations about vacations and movies. The length of the conversations ranged between five to seven minutes, depending on user responses.

The UTEP AGENT system [9], in which Paola Chat was implemented, interfaces with the Unity game engine to automate features during the interaction, i.e., generating dialogue, handling breathing and other gesture animations, cycle through the states of breathing, as well as recognizing user input during conversation .

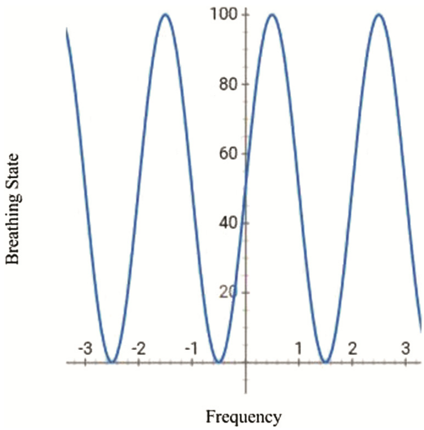


Fig. 3. Breathing oscillation function. The x-axis is the oscillation of sin(πx), where x is the frequency, and the y-axis is the change in the wave function for the values of the updating breathing state, with 0 the lowest point of exhale and 100 as the highest point of inhale.

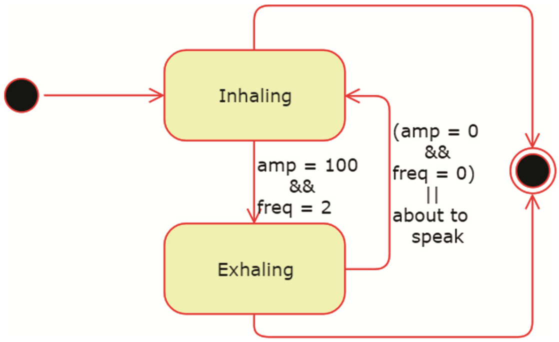


Fig. 4. The breathing cycles represented as a state diagram.

In the first conversation, Paola greeted the participant and began conversing on the topic of either vacations or movies; the order of the topics alternated as a part of the experimental design. Paola would occasionally ask questions where the participant's utterance would be either treated as a wildcard (where the content did not matter) or, based on keyword recognition, would trigger an appropriate response.

Table 1 shows examples of responses to questions asked by Paola and follow-up questions Paola asked during the interactions. For example, Paola would ask "*Have you seen any movies lately?*" If the participant responded that he or she had, then Paola would ask for more details about that movie. If, however, the participant responded no, then Paola would instead say "*It's okay! Tell me about your favorite movie, then. What is your favorite?*"

Table 1. Responses and follow-up questions.

Paola's utterance	*Example user response*	*Paola's follow-up*
Have you seen any movies lately?	Yes, I have	Oh, interesting. What was that movie about, was it good?
Have you seen any movies lately?	No, I haven't watched too many movies lately	Oh, I'm sorry to hear that. It's okay! Tell me about your favorite movie, then. Why is it your favorite?
Where would you go if you could visit any place in the world?	Probably the UK	Wonderful! If could visit only one place in Germany, it would be…
What is the one thing you would want to do for fun in your dream destination?	I guess, um, explore like Tokyo and the mountains of Japan, and shrines	Yeah, that's cool. Actually, I forgot to mention that one of my friends is going to…
In Beetlejuice, it's so funny when everyone starts singing the "Jump in the line song!" Do you know which song I'm talking about?	No	Yeah, come on, it goes something like'shake shake shake senora shake your body line…'
I didn't know [Jared Leto] was an actor until [Alexander the Great]. Have you seen that movie? it's like 12 years old	I probably watched it a while back, but I don't remember what that was about	I saw it a long time ago, so I don't remember much of it myself, but as I was saying I first knew him…

Empirical Evaluation. We used the Paola Chat application to evaluate users' perceptions of naturalness of the agent's breathing. The study was a within-subjects design in which one of the conversations had the ECA with the breathing animations and the other conversation had the ECA that did not use the breathing animations. The design was balanced for order of breathing/non-breathing and for order of conversation topic. After each conversation, participants were asked to complete a seven-point Likert-scale survey of naturalness, rapport, and social presence.

A total of 62 participants interacted with Paola. The population consisted of college students mostly aged 18–25 (about 85%; the remaining 15% were under age 30). The population consisted of 73% males. Further, 68% of the participants were native speakers of English. Of the participants, 21 identified as first-year college students, 11 as second-year college students, 13 as third-year college students, and 7 as fourth-year college students. The remaining 10 participants were in their fifth-year of study or above.

Before the interaction, each participant was asked to complete a demographic survey. Each also signed a permission to be video-recorded during the interaction. Participants were seated in front of a wall where Paola was projected (see Fig. 1). The two conversations each lasted about five to seven minutes, with the exact length of each interaction depending on the user's responses to Paola's questions.

After the first conversation, users were asked to complete a survey on the interaction. The interaction continued with a conversation on the other topic. The session would conclude with a final survey. Table 2 displays the 18-question survey participants completed; responses were entered on a 7-point Likert scale of users' perception of naturalness, rapport and of social presence, as used, for example, in [4, 10].

Table 2. Pre-interaction and post-interaction survey questions.

Q1	I feel that the agent understood me
Q2	The agent seemed disengaged
Q3	The agent was excited
Q4	The agent's movements were unnatural
Q5	The agent was friendly
Q6	The agent was not paying attention to me
Q7	The agent and I worked towards a common goal
Q8	The agent and I did not seem to connect
Q9	I sensed a physical connected with the agent
Q10	The agent's gestures were not lively
Q11	I feel the agent trusts me
Q12	I didn't understand the agent
Q13	I perceive that I am in the presence of another person in the room with me
Q14	I feel that the person is watching me and is aware of my presence
Q15	The thought that the person is not a real person crosses my mind often
Q16	The person appears to be sentient, conscious, and alive to me
Q17	I perceive the person as being only a computerized image, not as a real person
Q18	Overall, the agent's behavior seemed natural

3 Results

The study's results suggested that breathing did not affect the agent's perceived naturalness. Table 3 displays the average scores for rapport, naturalness, and social presence. Although the average rapport scores across the experimental and control conditions were normally distributed (Anderson-Darling test, $p = 033$ and $p = 0.26$, respectively), the absolute difference in average scores was small ($4.08 - 3.78 = 0.30$, on a scale from 1 to 7), and a t-test was not significant ($p = 0.76$). Similarly the t-tests for naturalness and social presence were also not significant (both $p = 0.69$).

Because the design of the experiment allowed for the participants to watch Paola as she spoke or listened, the design of Paola's utterances required an emphasis on the eliciting questions. It was important to elicit longer, more thoughtful responses than

Table 3. Naturalness, rapport, and social presence on both experimental and control conditions.

	Experimental condition		Control condition	
	Mean	StDev	Mean	StDev
Rapport	4.08	1.67	3.78	1.65
Naturalness	4.13	1.59	4.24	1.49
Social presence	3.14	1.67	3.23	1.76

simple affirmations or negations. Table 4 shows questions asked by Paola that elicited longer responses from the human participants.

Table 4. Responses to questions.

Paola's question	*User's response*
Can you think of any other artists that go back and forth with movies and music?	Uh, I think someone like Justin Bieber used to do that kind of stuff. Or, I can't think of any one too much. I'm not too good on artists and that kind of stuff
	Oh, Lady Gaga. She was in the American Horror Story, I thought she did really, good there
What sorts of new movie experiences do you think will come next?	Uh, VR or they sort of already do this, but like they move the seats around and so you can feel what's going on in the movie. You can feel it in your chair
	Um, I don't know
Tell me about the vacation [you last went on]	Um, so I've gone to Washington and that was a lot of fun, that was for an interview. I've also been to Vegas, that was a lot of fun. But I wasn't 21 so I couldn't gamble, but that was a lot of fun. I also went to Florida
	Oh, it was just winter break
What is the one thing you would want to do for fun in your dream destination?	Anything, as long as it's relaxing or so I could be at the beach. Um, I don't know, just relaxed, having fun basically
	I guess, um, explore like Tokyo and the mountains of Japan, and shrines
Have you seen [Alexander the Great]? It's like 12 years old	I probably watched it a while back, but I don't remember what that was about
	No
Can you think of your favorite scene of any movie?	I don't know, maybe the fight of Star Wars 3

In responding to questions about specifics, some users chose to response briefly, with short responses such as "Oh, it was just winter break," while others gave utterances over ten words, as shown in Table 4. One of Paola's questions, about things to do in dream destinations, generated longer utterances, but these utterances tended to be more general in tone, with fewer specifics. For example, users gave answers such as "relaxing" at or "exploring" their dream destinations.

As expected, some users responded to questions requiring specific knowledge (e.g., "*Can you think of any other artists that go back and forth with movies and music?*") with expanded statements, while other questions generated one-word responses. Responses to general questions (e.g., "*What is the one thing you would want to do for fun in your dream destination?*") generated more thoughtful responses from the users, and therefore utterances with a higher word count.

Inviting users to speak about their experiences or preferences produced longer utterances, too. In both topics, users responded to questions about their favorite movie scenes or dream vacations by responding with utterances longer than six words. This was also reflected in user responses to clarifying statements by Paola, for instance, when she asked "*Tell me about the vacation*" or "*What was that movie about?*"

4 Conclusion

When we first saw SimSensei, the PTSD-diagnosis agent, we noticed that it appeared that she was not breathing. This led to us to develop the Paola Chat application, which we then used for a perception study of naturalness of breathing in the agent. Our results suggested that breathing did not actually affect perceived naturalness, rapport, or presence.

Despite our expectation that increased naturalness from breathing would lead users to report greater rapport in the breathing condition than in the not-breathing condition, the study's results suggest that animation of breathing appears to neither increase nor decrease these perceptions. Of course, while breathing by an ECA does not increase naturalness, neither does it detract from naturalness. This suggests that ECAs using a breathing model similar to that of the Paola Chat agent can be at least as natural as a non-breathing agent. When combined with other features implemented, such as generating responses not only relevant to the topic but also relevant to the utterance to which Paola responds (see Table 1), the perceived social presence of ECAs could be increased.

So why did we perceive that SimSensei was not breathing, when the participants in our study did not notice the breathing in the Paola Chat agent? One of the differences between the two agents was the amount of dialog produced by the agent. SimSensei was mainly listening to the person with whom it interacted (except for some occasional questions, nods, and hand movements), while Paola was more conversational and contributed substantively to the conversation.

A second factor may be that the Paola Chat agent's breathing was displayed only visually and not auditorily. If someone is about to speak, you can sometimes hear the inhalation.

A limitation of this study is that the Paolo Chat application was fully automated rather than Wizard-of-Oz. That is, the application generated utterances as output to users and accepted responses as input independent of a human acting behind the scenes. The dialog models were developed beforehand, where responses could be either short responses recognized by key phrases (converging to a previously designed point of the conversation) or wildcard responses (which followed the flow of the conversation more generically while being able to handle longer utterances regardless of keywords). This technology, though, constrains the agent's conversational responsiveness.

Though Paola displayed animated gestures while talking, nodding, and now breathing, the application did not include more than nominal models of gaze and head movement. Adding these kinds of animations to an otherwise static embodied conversational agent might provide an even more humanlike appearance. These features could provide an improvement in perceived naturalness because breathing affects not only the torso and neck but shoulder movement and even timing of dialog generation.

References

1. Cassell, J. (ed.): Embodied Conversational Agents. The MIT Press, Cambridge (2000)
2. Pelachaud, C.: Studies on gesture expressivity for a virtual agent. Speech Commun. **51**(7), 630–639 (2009)
3. Huang, L., Morency, L.-P., Gratch, J.: Virtual rapport 2.0. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS (LNAI), vol. 6895, pp. 68–79. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23974-8_8
4. Gris, I., Novick, D., Camacho, A., Rivera, D.A., Gutierrez, M., Rayon, A.: Recorded speech, virtual environments, and the effectiveness of embodied conversational agents. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS (LNAI), vol. 8637, pp. 182–185. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-09767-1_22
5. Neff, M., Wang, Y., Abbott, R., Walker, M.: Evaluating the effect of gesture and language on personality perception in conversational agents. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS (LNAI), vol. 6356, pp. 222–235. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15892-6_24
6. Novick, D., Gris, I.: Building rapport between human and ECA: a pilot study. In: Kurosu, M. (ed.) HCI 2014. LNCS, vol. 8511, pp. 472–480. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07230-2_45
7. Rizzo, A.A., Scherer, S., DeVault, D., Gratch, J., Artstein, R., Hartholt, A., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Stratou, G.: Detection and computational analysis of psychological signals using a virtual human interviewing agent. In: Proceedings of the International Conference on Disability, Virtual Reality and Associated Technologies (2014)
8. Włodarczak, M., Heldner, M.: Respiratory turn-taking cues. In: Proceedings of Interspeech 2016, pp. 1275–1279 (2016)
9. Novick, D., Gris Sepulveda, I., Rivera, D.A., Camacho, A., Rayon, A., Gutierrez, M.: The UTEP AGENT system. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp. 383–384 (2015)
10. Novick, D., Gris, I., Camacho, A., Rayon, A., Gonzalez, T.: Bigger (Gesture) isn't always better. In: Kurosu, M. (ed.) HCI 2017. LNCS, vol. 10271, pp. 609–619. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58071-5_46

Quantifying Human Decision-Making: Implications for Bidirectional Communication in Human-Robot Teams

Kristin E. Schaefer[1](✉), Brandon S. Perelman[1], Ralph W. Brewer[1], Julia L. Wright[1], Nicholas Roy[2], and Derya Aksaray[3]

[1] US Army Research Laboratory, Aberdeen Proving Ground, Adelphi, MD 21005, USA
kristin.e.schaefer-lay.civ@mail.mil
[2] MIT, Cambridge, MA 02139, USA
[3] University of Minnesota, Minneapolis, MN 55455, USA

Abstract. A goal for future robotic technologies is to advance autonomy capabilities for independent and collaborative decision-making with human team members during complex operations. However, if human behavior does not match the robots' models or expectations, there can be a degradation in trust that can impede team performance and may only be mitigated through explicit communication. Therefore, the effectiveness of the team is contingent on the accuracy of the models of human behavior that can be informed by transparent bidirectional communication which are needed to develop common ground and a shared understanding. For this work, we are specifically characterizing human decision-making, especially in terms of the variability of decision-making, with the eventual goal of incorporating this model within a bidirectional communication system. Thirty participants completed an online game where they controlled a human avatar through a 14×14 grid room in order to move boxes to their target locations. Each level of the game increased in environmental complexity through the number of boxes. Two trials were completed to compare path planning for the condition of known versus unknown information. Path analysis techniques were used to quantify human decision-making as well as provide implications for bidirectional communication.

Keywords: Human-robot teaming · Bidirectional communication
Decision-making

1 Introduction

Human-agent teaming is a critical area of research because technological advancements are reaching the point where machines are able to make both independent and interdependent decisions. Due to these advancements, human team member roles are transitioning to more communication-based interactions supporting larger goals and intentions rather than direct control or teleoperation of the system [1]. One key limitation in the development of effective teaming has been the process of building a shared understanding of the mission space, whereby robotic team members can quickly and

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 361–379, 2018.
https://doi.org/10.1007/978-3-319-91581-4_27

accurately understand the human's intent and behaviors [2]. This is important because successful collaborative partnerships require communication, cooperation, and coordination between the acting members as they work towards a common goal [3], and distinct from prior work in supervisory control, which has often focused on ensuring the human has an accurate model of the robot's behavior. In a truly collaborative context, it is crucial that the robot also have situation awareness of the human partner.

1.1 Bidirectional Communication

Bidirectional communication is an area of current research that aims to improve development of common ground and shared understanding. It is especially critical forthe transformation of a robot from tool to team member [4] because it allows forjoint decision-making and development of shared mental model representations[5], and knowledge transfer through communication supports shared situationawareness [6, 7]. Increased autonomy, especially the capability for independentand interdependent decision-making in complex environments, supports this needfor the development of bidirectional communication. In order to effectivelycommunicate, it is important for both human and robotic team members to understand the decisions and decision-making processes within their team. The interpretation of an interaction or actions of a robotic team member can be directly influenced by the person's expectations for the interaction. Similarly, if the human teammate's actions and behaviors deviate from the robot's expectations, there will be a degradation in trust. Thus, if a robot team member can interpret and correctly predict the actions and behaviors of the human, then the robot can react accordingly. Bidirectional communication can also help both the team members understand when a decision may intuitively counter their own ideas or models by providing reasoning information that defines the appropriateness of its decisions, thus updating the team member's mental model and expectations for the task. In addition, the mode of communication and feedback capabilities have an effect on trust development in human-agent teams [8].

1.2 Human Decision-Making

While there are many types of decisions, this paper focuses on spatial decision-making. Human spatial decision-making is characterized by the ability to rapidly produce robust solutions to complex problems. For example, the Traveling Salesman Problem (TSP) requires participants to connect nodes, representing cities, to create the shortest tour among the nodes. While its instructions are simple, the TSP is NP-hard, and brute force solutions require calculating $(n - 1)!/2$ tours where n is the number of nodes. Despite the computational complexity, humans produce near-optimal solutions to this problem in linear time [9–12] using a combination of global and local spatial heuristics [13–15]. Due to the quality of the solutions and the speed at which they are produced, the decision-making mechanisms humans use to solve these problems are an area of study for the AI community, but the underlying mechanisms still remain unknown.

Naturalistic spatial decision-making tasks allow these mechanisms to be supported by guidance from top-down cognitive processing systems [16–19]. Humans are capable of adapting paths easily to mission requirements during naturalistic, real-world tasks, such as mission planning for unmanned aerial vehicles [20]. Yet, recent research demonstrated that aggregate human solutions tend to converge on not just one but several solution groups, each characterized by a distinct spatial mental model [21]. The adaptive nature of these top-down processes permits mission-dependent flexibility in the spatial decision-making process, and this characteristic has implications for bidirectional communication in human-agent teams.

Shared Mental Models. Understanding decision-making helps to classify the mental model for a task, which then guides expectations for interaction. Spatial mental models are mental representations of the environment [22–24], and weightings of the importance of features in those representations relative to goals [22]. Spatial mental models directly impact a solution to a given spatial decision-making problem, as well as their evaluations of solutions generated by other humans and algorithms. This has direct implications for all manner of human-agent teaming problems. For example, in collaborative spatial decision-making, an algorithm may propose a route to a human who can either accept or choose to replan it. This replanning or retasking degrades performance and situation awareness, and increases workload [25]. In addition, divergence between the human team members' spatial mental models and the actions taken by an intelligent agent can reduce predictability and degrade trust [26–28]. Conversely, spatial mental models that are similar to an agent's suggestion can improve agent trust, and increase the rate of acceptance for that solution [21]. This is especially true for cases where agents are unable to articulate their reasoning for producing solutions that may contradict human teammates' spatial mental models. Therefore, this area represents a potential target for future research in bidirectional communication for the purpose of achieving consensus between human spatial mental models and intelligent agent problem solving mechanisms.

Implications for Agent Development and Teaming. Knowing how humans make decisions could help a robot to derive a model of the team member's planning model, which allows the robot to infer future human behavior, and provides the needed context to communicate its state and goals in the same representation as the human. Such an extrapolation could greatly reduce the need for explicit communication (e.g., it might suffice for a robot to observe nodding or a hand que to infer how human will act next). Moreover, matching both representation and goals could result in an efficient search over the action space (e.g., a robot could narrow its search space based on expected human behavior). For intractable problems, knowing the optimal solution may not be possible. In that case, knowing how humans solve a problem could be a benchmark when developing robot algorithms. Further, understanding the limitations of a human can help teaming in such a way that the robot can take the initiative of being the main actor (e.g., computing a plan) in cases where the human is limited.

1.3 Current Work

The main objective of this research is to understand similarities and differences among human spatial decision-making processes as they apply to future human-agent teaming. When developing new spatial planning algorithms for robotic systems that will be collaborating with people to complete a task (e.g., moving objects around a room), it is important to characterize and compare the behavior of each of the agents under different conditions. Further, since each agent applies its own spatial mental model or algorithm to solve a given problem, in order to achieve robust collaboration and teamwork it is critical to recognize how the decision-making processes of each agent will handle increasing environmental complexity and uncertainty. Where disparities exist between the resultant robot and human behaviors, bidirectional communication can be used as a means to achieve an optimal solution collaboratively. A first step toward achieving this goal is to characterize human spatial decision-making behaviors in the proposed tasks. For this study, an online game was developed to assess human spatial decision-making processes involved with controlling a virtual avatar through a virtual room with the purpose of pushing virtual boxes from a set of start locations to a set of end locations. The design of the study was such that each level represented an increase in environmental complexity, and the two conditions represented an increase in task difficulty based on the availability of planning information.

2 Methodology

2.1 Participants

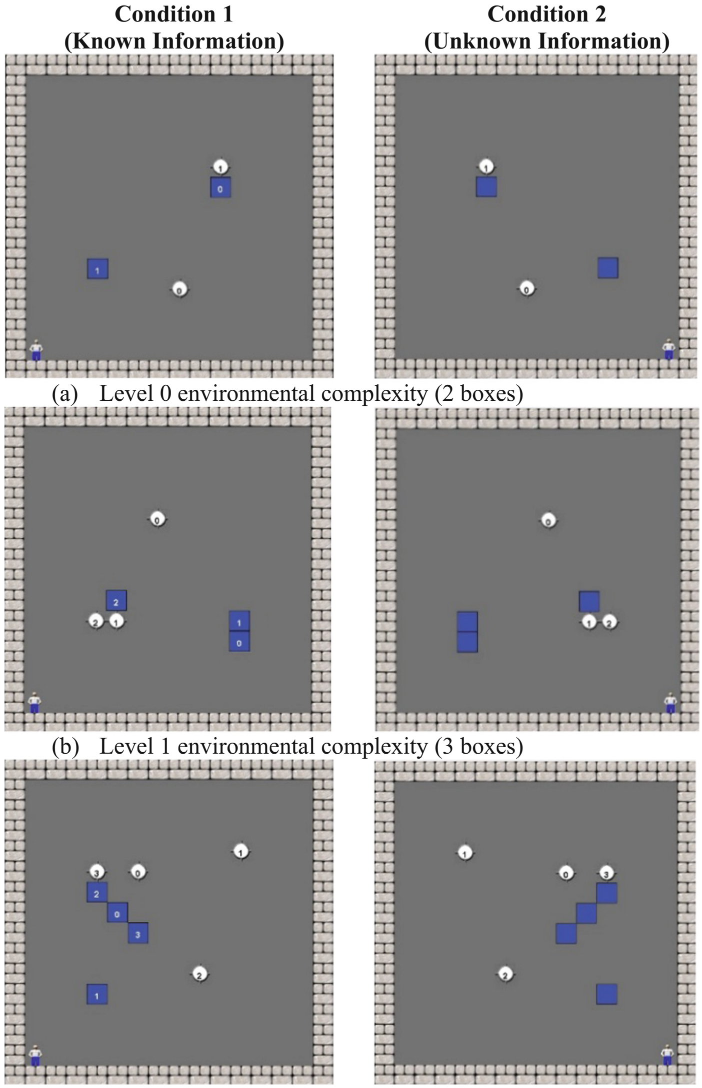
Thirty participants between the ages of 18 and 60 were recruited. This age restriction was selected to reduce variance in participants' spatial abilities. In prior studies involving tasks requiring spatial working memory, age-related cognitive decline reduces navigation speed [29] and overall task performance [30–32].

2.2 Game Development and Task

A Java Applet was built around similar game dynamics as the puzzle game Sokoban [33]. Sokoban, developed by Thinking Rabbit game studio in 1982, is a logic puzzle designed for the user to push objects (stones, boxes, etc.) around a playing field to a goal area in the fewest moves possible. For our study, the main game space for all levels was a 14×14 square grid surrounded by a brick wall on all sides. The grid space was developed to match the laboratory facilities at MIT to allow for future comparison of human and robot decision-making. The difference in design between this application and the original Sokoban game was that typically levels only had minimal number of solutions, while the open area of this playing field made for exponentially more paths to reach a solution.

There were nine levels (Level 0 through Level 8) that increased in the number of boxes from two to 10 boxes (Fig. 1). The number and location of the boxes and target locations were devised in such a way to represent increasing environmental complexity. In order to investigate the variability in human decision-making, different patterns, clusters, and spatially distant blocks were used while choosing the initial and target locations of the blocks. The game levels were designed in such a way that the optimal (or very close to optimal) solution was not obvious. This helps to identify variability in human decision-making behaviors.

The start location of the avatar (represented by a person) was always in the same start corner. The avatar could only move up, down, left, or right (no diagonals) and could only push (not pull) the boxes. Therefore, the initial placement of the boxes were located a certain distance from the boundaries to ensure the existence of a solution. In order to avoid infeasibility (e.g., deadlock), the exterior 2 cells were intentionally left

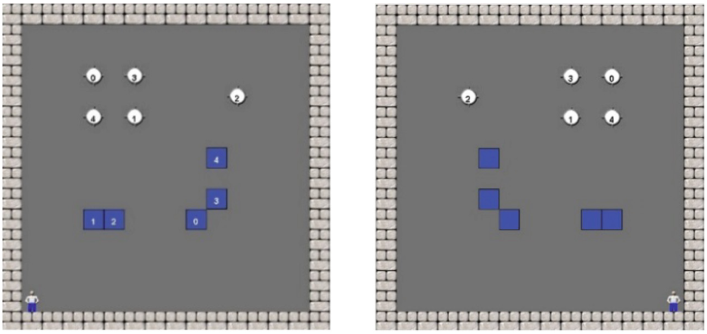


(a) Level 0 environmental complexity (2 boxes)

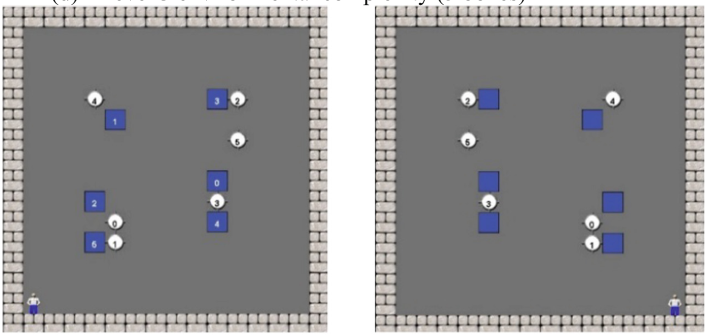
(b) Level 1 environmental complexity (3 boxes)

(c) Level 2 environmental complexity (4 boxes)

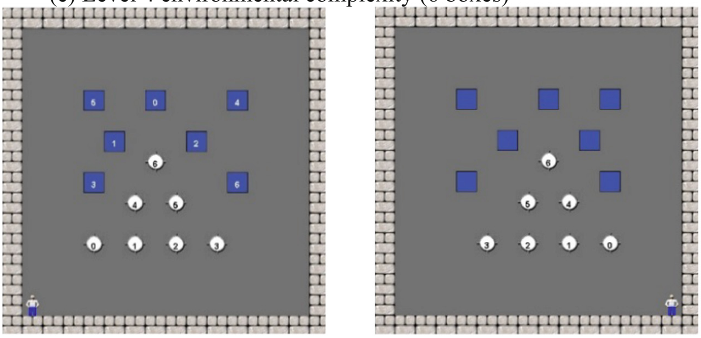
Fig. 1. Each game level (a)–(h) represents increasing environmental complexity. For Condition 2 (Unknown Planning Information), the entire game board was transposed over the y-axis making each condition directly comparable but unique.



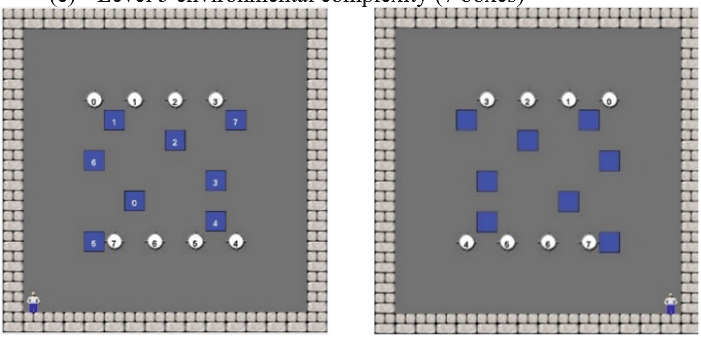
(d) Level 3 environmental complexity (5 boxes)



(e) Level 4 environmental complexity (6 boxes)

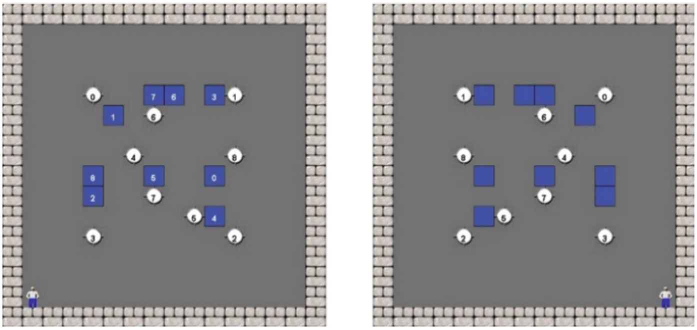


(e) Level 5 environmental complexity (7 boxes)

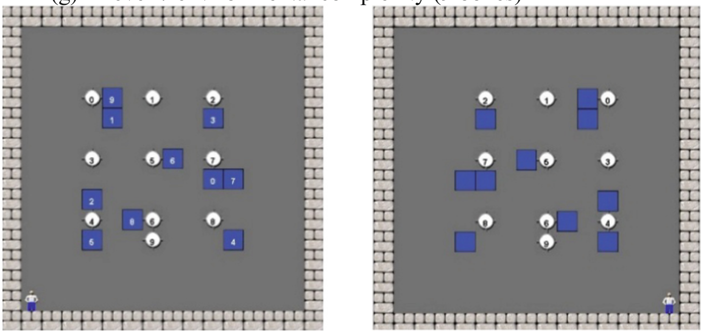


(f) Level 6 environmental complexity (8 boxes)

Fig. 1. (*continued*)



(g) Level 7 environmental complexity (9 boxes)



(h) Level 8 environmental complexity (10 boxes)

Fig. 1. (*continued*)

blank and boxes were not placed in corridor like shapes. An undo option to backup through previous moves, as well as a reset level option were available so that it was always possible to reach a solution.

The overall goal was to move all boxes from their initial locations to their target locations. To this end, there were two main criteria to determine the overall trajectory, the sequence the boxes should be moved and calculating the shortest path from a box's initial location to its target location. Participants completed two conditions representing increasing task difficulty. For Condition 1 (Known Planning Information), all boxes and target locations were known, such that participants had to control the human avatar through the virtual environment and push Box 1 to Target Location 1. Participants were instructed that the numbering on the boxes and target locations were only there to inform which box was connected with which target location. They could complete the task in any order. For Condition 2 (Unknown Planning Information), the boxes were unlabeled however all the target locations were labeled. The box numbers only became visible once the human avatar pushed the box to a new grid square location.

2.3 Design

The experimental design was a 2 condition (known versus unknown planning information) × 9 environmental complexity (game levels ranging from 2–10 boxes) within-subjects design. Participants completed Condition 1, Levels 0 through 8, followed by Condition 2, Levels 0 through 8. There were two main hypotheses.

Hypothesis 1: Overall path length, number of movements, and time to completion will be longer when the environment is more complex (i.e., more boxes) and when the task is more difficult (i.e., changes in the amount of previously known information about the task).

Hypothesis 2: Based on our prior research, there will be more than one "human" way of making decisions to plan a path and move the boxes. The Algorithm for finding the Least Cost Areal Mapping between Paths (ALCAMP) [34] will be used to quantify the divergence among participants' solutions in Condition 1 and Condition 2 for each level. Higher levels of environmental complexity associated with the levels will produce greater divergence (i.e., variability) in solutions, and we expect that the availability of planning information (manipulated in each Condition) will also impact the divergence among solutions in each level.

2.4 Analysis

Performance. Specific decision-making time and movements were recorded for all levels across both trials. Decision-making times included planning time (i.e., time to first movement), total completion time (i.e., time till last box was placed on the correct target), and action time (i.e., total completion time minus planning time). The actual decisions were analyzed by looking at the total number of moves to complete each level, and nearest-neighbor analysis. The nearest-neighbor analysis calculated the number of participants who first moved to the closest box to the start location compared to another box. This analysis provides insight into whether or not they used a local versus global strategy.

Variability. A novel approach was used to characterize the variability among participants' solutions for each of the spatial problems in each level. The variability among solutions is important to understand differences between individuals' spatial solutions and to determine predictability in decision-making behaviors. Thus, increasing solution variability corresponds to decreasing predictability for both human- and algorithm-produced solutions. To measure the variability, participants' solutions were pooled within each level representing environmental complexity and condition of task difficulty. ALCAMP [34] was used to compare all solutions within each pool in a pairwise-exhaustive fashion. The resultant values of this analysis reflect the divergence between the pair of paths as measured by Euclidian divergence among the grid squares, such that large values indicate that the two solutions are different and small values indicate that they are similar. Finally, these values were used to populate a symmetric dissimilarity matrix through a distance matrix. The mean of the upper or lower triangle indicates the average dissimilarity among all of the solutions for a given level and condition – the higher the value, the greater the variability among participants' solutions to that problem. This technique has been used in previous research to infer consensus in spatial decision-making processes [11, 21].

3 Results

3.1 Performance Analysis

Total Completion Time. A 2 condition (known vs unknown boxes) × 9 levels of environmental complexity (levels 0–8) repeated measures within-subjects ANOVA was conducted to assess total completion time. There was a main effect of condition, $F(1,24) = 15.98$, $p = .001$, $d = 1.03$, where Condition 1 (known planning information) was longer, $M = 49.894$ s, SE = 2.553, than Condition 2 (unknown planning information), $M = 44.049$ s, SE = 1.898. There was a main effect of level, $F(8, 17) = 78.89$, $p < .001$, $d = 1.35$, whereby increased environmental complexity led to increased completion time, and an interaction, $F(8, 17) = 8.821$, $p < .001$, $d = 0.41$, see Fig. 2. These results show that increasing the amount of information available to an agent can increase processing times despite providing important cues to objects in the environment.

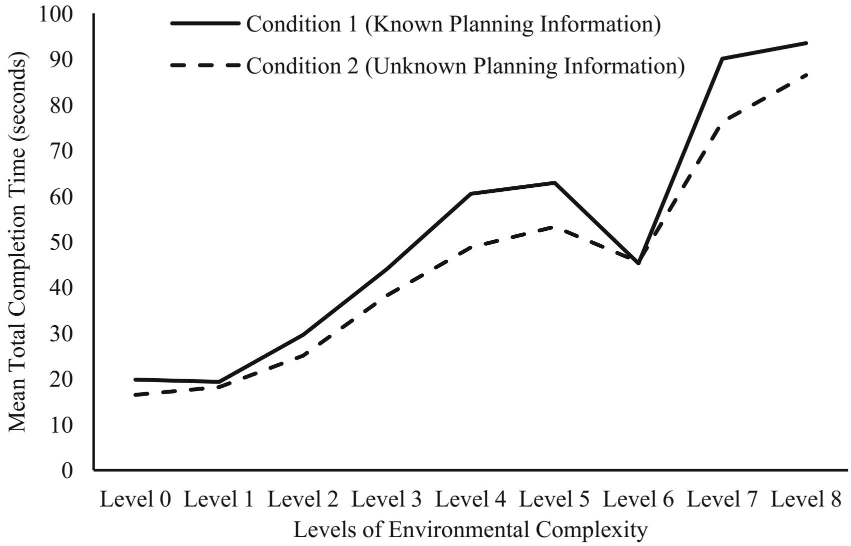


Fig. 2. Total completion time (seconds) on each game level representing environmental complexity (game level) for the known and unknown planning information conditions

Paired samples *t*-tests were conducted for each level of environmental complexity to compare total completion time in the known and unknown planning information conditions. There was a significant difference in scores for Levels 0 ($p = .020$), 2 ($p = .008$), 3 ($p = .009$), 4 ($p < .001$), 5 ($p = .004$), and 7 ($p < .001$), whereby completion time was significantly longer for Condition 1 (known planning information) than Condition 2 (unknown planning information). Results are reported in Table 1. These results show that the effect between the two conditions collapsed completely on Level 6, and partially on Level 8, likely owing to characteristics of those specific environments.

Table 1. Paired samples *t*-tests for total completion time

Game level	Mean time 1	SD 1	Mean time 2	SD 2	*t*	*p*	*d*
0	19.851	6.827	16.513	5.294	$t(26) = 2.479$	.020	0.48
1	19.342	6.401	18.211	4.015	$t(27) = 1.272$	.214	0.24
2	29.614	10.303	25.077	5.351	$t(27) = 2.851$	.008	0.54
3	44.052	11.258	38.273	7.160	$t(27) = 2.830$	.009	0.53
4	60.510	16.714	48.775	1.4152	$t(27) = 4.373$	<.001	0.83
5	62.897	22.463	53.272	13.360	$t(26) = 3.117$	.004	0.60
6	45.275	11.520	45.611	12.612	$t(26) = -.139$	.891	–0.03
7	90.077	26.434	76.256	20.313	$t(27) = 5.855$	<.001	1.11
8	93.486	29.564	86.447	27.756	$t(28) = 1.893$	.069	0.35

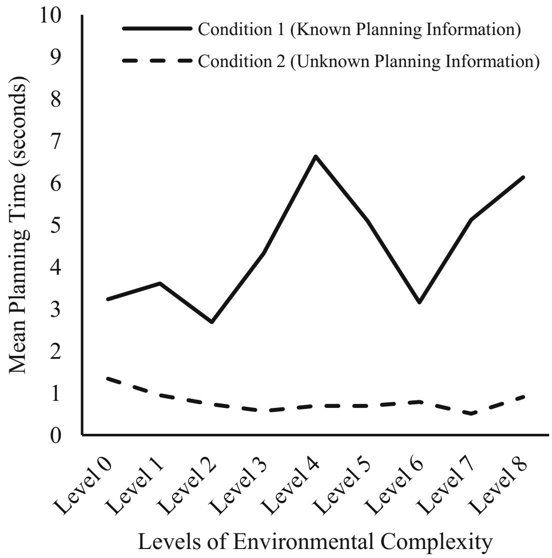
Note. Time is in seconds

Planning and Action Time. In order to determine whether the source of the total completion time effects described above were due to differences in planning or action, we split the total completion time into a planning phase (duration between trial presentation and the first move) and action time (the remainder of the total completion time – planning time). A 2 condition × 9 levels of environmental complexity repeated measures within-subjects ANOVA was conducted to assess planning time. There was a main effect of condition, $F(1,15) = 24.33$, $p < .001$, $d = 2.006$, where Condition 1 (known planning information) was longer, $M = 3.123$ s, SE = 0.501, than Condition 2 (unknown planning information), $M = 1.375$ s, SE = 0.129. There was a marginal main effect of level, $F(8, 120) = 1.967$, $p = .056$, $d = 0.028$, whereby increased environmental complexity led to increased planning time, and an interaction, $F(8, 120) = 2.776$, $p = .007$, $d = .038$. Paired samples *t*-tests were conducted for each level of environmental complexity to compare planning time in the known and unknown planning information conditions. There was a significant difference in scores for Levels 1 ($p = .005$), and Levels 2 through 8 ($p < .001$), whereby planning time was significantly longer for Condition 1 (known planning information) than Condition 2 (unknown planning information). Results are reported in Table 2.

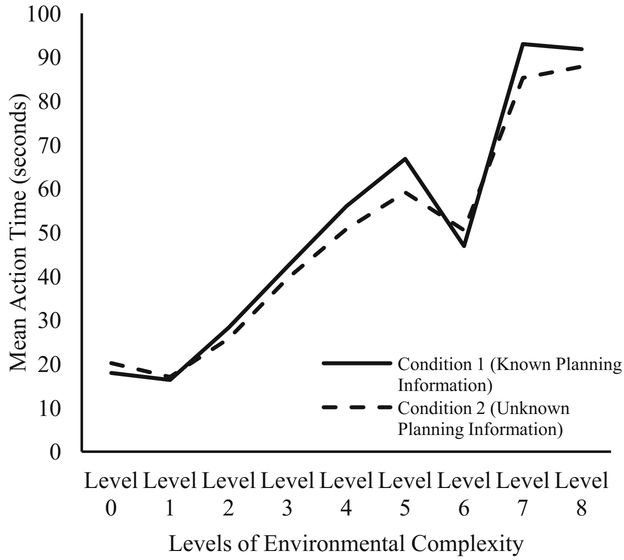
Table 2. Paired samples *t*-tests for planning time

Game level	Mean time 1	SD 1	Mean time 2	SD 2	*t*	*p*	*d*
0	3.230	1.281	2.765	1.341	$t(24) = 1.591$	.125	0.32
1	3.608	3.213	1.891	0.948	$t(25) = 3.077$	.005	0.60
2	2.686	1.326	1.369	0.732	$t(24) = 5.548$	<.001	1.11
3	4.326	3.542	1.268	0.574	$t(25) = 4.627$	<.001	0.91
4	6.629	6.468	1.436	0.694	$t(28) = 4.452$	<.001	0.83
5	5.099	4.601	1.496	0.694	$t(27) = 4.405$	<.001	0.83
6	3.155	2.036	1.453	0.790	$t(25) = 4.563$	<.001	0.89
7	5.124	3.815	1.211	0.509	$t(24) = 5.227$	<.001	1.05
8	6.134	5.246	1.713	0.907	$t(26) = 4.542$	<.001	0.87

Note. Time is in seconds



(a)



(b)

Fig. 3. Mean task times across levels of environmental complexity for planning time (a) and action time (b) where the solid black line represents Condition 1 (Known Planning Information) and the dashed line represents Condition 2 (Unknown Planning Information)

These results show that, while planning time generally increased with environmental complexity when boxes were known (thought the layout of the environment clearly played a role as well, as shown by the dip in planning time for Level 6), planning time essentially dropped to floor when boxes were unknown. One interpretation of this result is that participants adopted a very simple local decision-making strategy when the box numbers were unknown, as opposed to the global search performed when all planning information was presented. In addition, the paired samples *t*-tests showed that the effect of uncertainty on planning requires a minimum amount of environmental complexity (i.e., number of boxes) to manifest.

To assess action time, a 2 condition × 9 levels of environmental complexity repeated measures within-subjects ANOVA was conducted to assess action time. There was a main effect of condition, $F(1,29) = 5.683$, $p = .024$, $d = .359$, where Condition 1 (known planning information) was longer, $M = 51.040$ s, SE = 4.202, than Condition 2 (unknown planning information), $M = 48.464$ s, SE = 4.172. There was a main effect of level, $F(8, 232) = 99.828$, $p < .001$, $d = .365$, whereby increased environmental complexity led to increased completion time. There was not a significant interaction, $p = .088$. Figure 3 depicts the mean planning times and mean action times for both conditions across all levels of environmental complexity.

Paired samples *t*-tests were conducted for each level of environmental complexity to compare action time in the known and unknown planning information conditions. There was a significant difference in scores for Level 4, $t(29) = 2.217$, $p = .035$, $d = 0.40$; Level 5, $t(29) = 2.045$, $p = .050$, $d = 0.37$, Level 6, $t(29) = -2.116$, $p = .043$, $d = -0.39$, and Level 7, $t(29) = 2.291$, $p = .029$, $d = 0.04$. These results show that the interaction effect between information availability and environmental complexity nearly disappears when removing the variance attributable to the planning phase of problem solving. Thus, these results taken together show that the differences reflect differences in planning for spatial problem solving rather than the action of actually moving the avatar to solve the problem.

Number of Moves. A 2 condition × 9 levels of environmental complexity repeated measures within subjects ANOVA was conducted to assess number of moves needed to complete the task. There was a marginal main effect of condition, $F(1,21) = 3.85$, $p = .063$, d = .337, where Condition 2 (unknown planning information) required more moves, M = 117.64, SE = 1.595, than Condition 1 (known planning information), M = 115, SE = 1.590. There was a main effect of level, $F(1, 21) = 2707.23$, $p < .001$, d = 2.48, whereby increased environmental complexity led to increased completion time. There was significant interaction, $F(1, 21) = 4.33$, $p = .050$, d = .042. Paired samples *t*-tests were conducted for each level of environmental complexity to compare total number of moves in the known and unknown planning information conditions. There was a significant difference in scores for Level 1, $t(26) = -3.389$, $p = .002$, $d = -.65$; and a marginal significant difference in scores for Level 8, $t(28) = -2.028$, $p = .052$, $d = -.38$. The result of this analysis indicate that the conditions did not have a meaningful impact on the number of moves required to complete each problem. Rather, the number of moves was mostly impacted by the environmental complexity.

Nearest Neighbor Analysis. In order to test the hypothesis that participants relied more heavily on local decision-making heuristics when boxes were unknown, we calculated the number of participants who first moved the box closest to the starting position (*nearest neighbor preference*), at each level of environmental complexity for each trial condition. A high percentage of participants exhibiting nearest neighbor preference would indicate a very simple local decision-making strategy. A lower percentage would indicate that participants employed a more global strategy. Figure 4 shows the percentage of participants who interacted with the closest box first for each level, in both experimental condition.

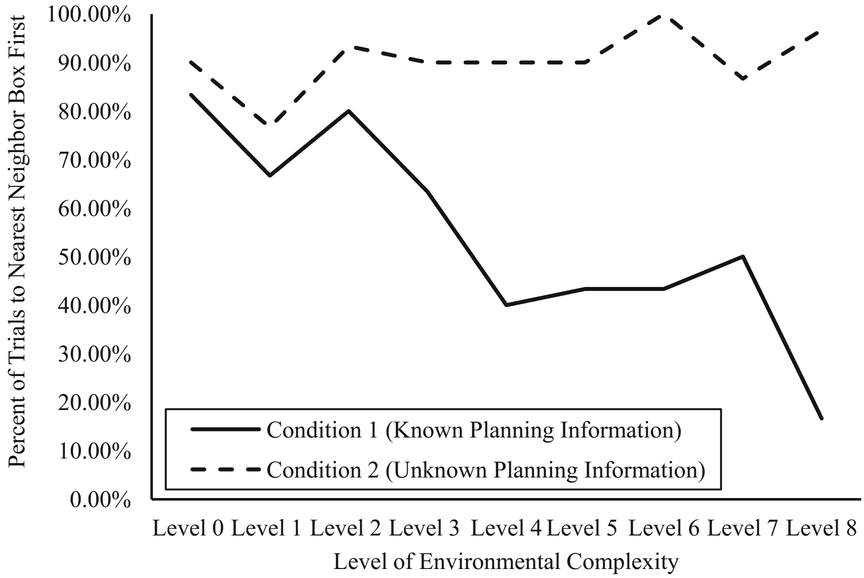


Fig. 4. Nearest neighbor analysis indicates the percentage of trials on which participants visited the box closes to the starting location first.

Participants' nearest neighbor preference was near-ceiling when the boxes were unknown. When the boxes were known in advance, the percentage of participants showing a nearest-neighbor preference decreased with increasing environmental complexity. These results, taken together, substantiate the hypothesis that participants typically employed global decision-making strategies when the box identities were known, but resorted to local decision-making strategies in the absence of that information.

3.2 Variability Analysis

Using the aforementioned procedure for calculating the average divergence in each condition and level, we see that variability in participants' solutions increases roughly linearly with increasing environmental complexity, once complexity reaches a certain threshold (in this case, it appears to be 5 boxes). A 2 condition (Known vs. Unknown planning information) $\times$ 9 environmental complexity (Levels 0–8) ANOVA revealed significant main effects of both Conditions, $F(1, 7812) = 227.33$, $p < .001$, and Level, $F(8, 7812) = 2305.92$, $p < .001$, as well as an interaction effect between these two variables, $F(8, 7812) = 28.27$, $p < .001$ (see Fig. 5). Post hoc analysis using Tukey's HSD tests showed significant effects between Conditions for levels 0 ($p = .022$), 2 ($p < .001$), 4 ($p < .001$), 5 ($p < .001$), 7 ($p < .001$) and 8 ($p < .001$). In all of these

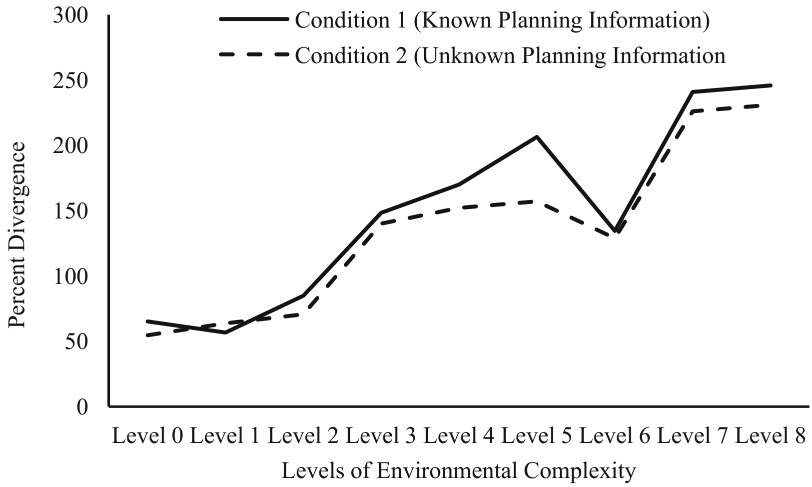


Fig. 5. The divergence value indicates the number of grid points by which solutions differed or variance in solutions. The average divergence among participants' solutions show very little differences in solutions for lower levels of environmental complexity (Levels 0–2) but a large increase in the number of possible solutions starting at Level 3.

cases, participants' solutions exhibited greater average divergence when the boxes were known (i.e., Condition 1) versus unknown (Condition 2). Note that the dip in divergence in both conditions on Level 5 (7 boxes) likely reflects characteristics of that particular environment.

4 Discussion

Communication is described by the research community as a reciprocal process where teammates send and receive information that form and reform the team's attitudes, behaviors, and cognition [35] whereby a shared body of knowledge can be used to develop shared expectations, allowing for improved team performance without explicit coordination [36, 37]. In the past, a variety of methods to facilitate bidirectional communication have been explored. For example, transparent user displays can convey agent intent [2], as well as its goals, reasoning, and projected outcomes [38–40]. While a multimodal approach to communication can reduce workload and degraded situation awareness [41] by using both implicit (nonverbal – behaviors, actions) and explicit (voice, natural language or auditory) communication modalities [42]. Considerable efforts have gone into determining the type, amount, modality, and rate by which information should be communicated between team members (e.g., Situation Awareness agent-based Transparency Model [40]). But perhaps key to the development of effective bidirectional communication within human-agent teams is the need for team members to be aware of goals, reasoning, actions, and projected outcomes of their teammates [43–46]. Therefore, a major step to developing appropriate bidirectional communication is being able to quantify human behavior across tasks. If human behavior does not match the robots' models or expectations, there can be a degradation in trust that can impede team performance and may only be mitigated through explicit communication.

4.1 General Discussion

This was the first study in a set of studies using this paradigm. It was designed to advance the technical capabilities of a robot to more accurately perceive and interpret human team member behavior, and to develop appropriate bidirectional communication required for future collaborative tasking. By first looking at quantifying human behavior, we can provide a foundation for understanding how human expectations for planning and spatial task solutions are formed. This is essential for future teaming because when human expectations do not match robot behaviors then degradations in trust can occur. Therefore, quantifying the decision space can provide insights into identifying when and how bidirectional communication could mitigate divergences in human and robot team behaviors.

Human Performance. The results of the present study showed that completion times generally increased with increasing environmental complexity. Furthermore, participants generally took longer to complete the levels when the box contents were known. Separating participants' solution times into planning times and action times showed that the majority of this discrepancy between the two information availability conditions was due to differences in time spent in planning rather than action. When the box numbers were visible to participants, participants took longer to begin moving than when the box numbers were not known, and we believe this time was spent analyzing the environment and planning their moves. Curiously, this increase in planning time did not translate to increased efficiency, as participants' solutions did not vary between the two information availability conditions in terms of the number of moves. Generally, this result indicates that perfect world knowledge did not improve performance, and actually reduced the speed with which participants completed each level. Understanding variance in planning and completion times can provide insights into situations that may require more explicit communication between team members to clarify the underlying reasoning process for the decision being made, as well as help to determine timing associated with providing feedback to a team member.

Global versus Local Decision-Making and Implications for Bidirectionality. In order to quantify decision-making behaviors, as well as further investigate the planning time difference described above, we performed a simple analysis to determine whether participants were using a local decision-making heuristic - nearest neighbor. The nearest neighbor analysis showed that nearly all participants employed a nearest neighbor heuristic when the box numbers were unknown, visiting the closest box first. When box numbers were known, participants appeared to increasingly leverage global decision-making strategies. This result, taken in the context of the performance results, shows that perfect world knowledge, which facilitates global decision-making processes, does not produce any marked advantage in efficiency or speed over simple local decision-making heuristics for these problems. This is important for bidirectional communication, as it shows that more complex decision-making algorithms may produce only marginal performance gains over simpler algorithms, at a cost of being far more difficult to explain to human teammates and the computational complexity of the algorithm itself.

Predictability of Decisions. An important part of human-agent teams is the extent to which agents can predict one another's actions. This can be viewed as a function of the number of different solutions that a group of agents will produce, or that a stochastic algorithm will produce on successive runs. Greater differences among solutions indicates that those solutions will be harder for teammates to predict, whereas if all teammates' solutions converge to only a few possibilities it will be easier to predict their actions. In order to examine the predictability of human solutions, we calculated the mean pairwise divergence among all solutions to each of the problems, in each condition. The main effect of environmental complexity was characterized by a general increase in divergence with increasing environmental complexity, once the complexity of the environment increased beyond a threshold; in the present study, the threshold was five boxes. This means that most humans will make similar solutions when environmental complexity is low suggesting that additional explicit communication may not be needed since the likelihood that expectations will match behaviors is high. However, when the environmental complexity reaches a set level, the number of possible solutions and variance between those solutions greatly increases leading to more unpredictable human behavior. These results also showed that with the exception of Level 6, after this threshold participants solutions diverged more when the box numbers were not shown. One interpretation of this finding, in light of the previous results, is that participants' reliance on local decision-making heuristics when the box numbers were shown reduced the variance across their solutions.

Summary Discussion. The results described above, taken together, show that local decision-making heuristics are sufficient for this task, as global processing takes longer, does not improve performance or efficiency, and increases the divergence among participants' solutions. These solutions would thus be harder for teammates to predict. Beyond a certain level of environmental complexity, bidirectional communication becomes increasingly important because the range of possible solutions to a given problem increases substantially. In these cases, bidirectional communication will be necessary to promote shared situation awareness and trust, and to facilitate fluid, flexible interaction between humans and non-human intelligent agents.

4.2 Implications on Algorithm Development

Bidirectional communication has various impacts on the development of algorithms for the robot decision-making in collaborative missions. Specifically, shared understanding of the mission goals and mental models could minimize uncertainty in team decision making, and result in more predictable consequences. From the algorithmic perspective, predictable results after taking actions would greatly reduce online computations such as replanning because less deviations from the original plan would be observed. Also, understanding human decision-making and limitations could help developing robots that can autonomously decide when and how to help humans. For example, robot might explicitly offer help when human spends a lot of time for planning the move, or robot could infer a specific part of the problem (e.g., furthest area from the human) and start working on that to shrink the decision problem of human. Overall, both implicit (e.g., posture, gesture) or explicit (e.g., natural language, feedback through displays) communication play an important role when developing decision-making strategies for robots that are expected to operate with humans in complex missions.

Acknowledgment. Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Scholtz, J.: Theory and evaluation of human robot interactions. In: 36th Annual Proceedings of the International Conference on System Sciences. IEEE, Hawaii (2003)
2. Schaefer, K.E., Straub, E.R., Chen, J.Y.C., Putney, J., Evans, A.W.: Communicating intent to develop shared situation awareness and engender trust in human-agent teams. Cogn. Syst. Res. **46**, 26–39 (2017)
3. Groom, V., Nass, C.: Can robots be teammates? Benchmarks in human-robot teams. Interact. Stud. **8**(3), 483–500 (2007)
4. Oh, J., et al.: Integrated intelligence for human-robot teams. In: Kulić, D., Nakamura, Y., Khatib, O., Venture, G. (eds.) ISER 2016. SPAR, vol. 1, pp. 309–322. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-50115-4_28
5. Kaupp, T., Makarenko, A., Durrant-Whyte, H.: Human-robot communication for collaborative decision making – A probabilistic approach. Robot. Auton. Syst. **58**(5), 444–456 (2010)
6. Lyons, J.B.: Being transparent about transparency: a model for human-robot interaction. In: Proceedings of the AAAI Spring Symposium, Palo Alto, CA. AAAI (2013)
7. Sycara, K., Sukthankar, G.: Literature Review of Teamwork Models (Report No. CMU-RI-TR-06-50). Carnegie Mellon University, Pittsburgh (2006)
8. Schaefer, K.E., Chen, J.Y.C., Szalma, J.L., Hancock, P.A.: A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Hum. Factors **58**(3), 377–400 (2016)
9. Dry, M., Lee, M.D., Vickers, D., Hughes, P.: Human performance on visually presented traveling salesperson problems with varying numbers of nodes. J. Probl. Solving **1**(1), 20–32 (2006)
10. MacGregor, J.N., Ormerod, T.: Human performance on the traveling salesman problem. Percept. Psychophys. **58**(4), 527–539 (1996)
11. Perelman, B.S., Mueller, S.T.: Identifying mental models of search in a simulated flight task using a pathmapping approach. In: Proceedings of the 18th International Symposium on Aviation Psychology (2015)
12. Pizlo, Z., Stefanov, E., Saalweachter, J., et al.: Traveling salesman problem: a foveating pyramid model. J. Probl. Solving **1**(8), 83–101 (2006)

13. Best, B.J., Simon, H.A.: Simulating human performance on the traveling salesman problem. In: Proceedings of the 3rd International Conference on Cognitive Modeling, Groningen, Netherlands, pp. 42–49. ICCM (2000)
14. Perelman, B.S., Mueller, S.T.: Considerations influencing human TSP solutions and modeling implications. In: Reitter, D., Ritter, F.E. (eds.) Proceedings of the 14th International Conference on Cognitive Modeling. University Park, PA (2016)
15. Kong, X., Schunn, C.D.: Information seeking in complex problem solving. In: Proceedings of the 8th International Conference on Cognitive Modeling, Oxford, UK, pp. 261–266. ICCM (2007)
16. Blum, A., Chawla, S., Karger, D.R., et al.: Approximation algorithms for orienteering and discounted-reward TSP. SIAM J. Comput. **37**, 653–670 (2007)
17. Ragni, M., Wiener, J. M.: Constraints, inferences and the shortest path: Which paths do we prefer? In: Proceedings of the Annual Meeting of the Cognitive Science Society, pp. 2216–2221. CogSci, Sapporo (2012)
18. Tenbrink, T., Seifert, I.: Conceptual layers and strategies in our tour planning. Cogn. Process. **12**, 109–125 (2011)
19. Tenbrink, T., Wiener, J.: The verbalization of multiple strategies in a variant of the traveling salesman problem. Cogn. Process. **10**, 143–161 (2009)
20. Perelman, B.S., Mueller, S.T.: Examining memory for search using a simulated aerial search and rescue task. In: Proceedings of the 17th International Symposium on Aviation Psychology, Dayton, OH (2013)
21. Perelman, B.S., Evans III, A.W., Schaefer, K.E.: Mental model consensus and shifts during navigation system-assisted route planning. In: Proceedings of the Human Factors and Ergonomics Society, 61(1), 1183–1187. HFES, Austin (2017)
22. Brunyé, T.T., Taylor, H.A.: Extended experience benefits spatial mental model development with route but not survey descriptions. Acta Psychol. **127**, 340–354 (2008)
23. Jahn, G., Johnson-Laird, P.N., Knauff, M.: Reasoning about consistency with spatial mental models: hidden and obvious indeterminacy in spatial descriptions. In: Freksa, C., Knauff, M., Krieg-Brückner, B., Nebel, B., Barkowsky, T. (eds.) Spatial Cognition 2004. LNCS (LNAI), vol. 3343, pp. 165–180. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32255-9_10
24. Tversky, B.: Spatial mental models. Psychol. Learn. Motiv. **27**, 109–145 (1991)
25. Clare, A.S., Maere, P.C.P., Cummings, M.L.: Assessing operator strategies for real-time replanning of multiple unmanned vehicles. Intell. Decis. Technol. **6**, 221–231 (2012)
26. Chen, J.Y.C., Barnes, M.J.: Human-agent teaming for multirobot control: a review of human factors issues. IEEE Trans. Hum.-Mach. Syst. **44**(1), 13–29 (2014)
27. Hancock, P.A., Billings, D.R., Schaefer, K.E.: Can you trust your robot? Ergon. Des. **19**(3), 24–29 (2011)
28. Ososky, S., Schuster, D., Phillips, E., Jentsch, F.G.: Building appropriate trust in human-robot teams. In: Proceedings of the AAAI Spring Symposium. AAAI, Stanford (2013)
29. Zakzanis, K.K., Quintin, G., Graham, S.J., Mraz, R.: Age and dementia related differences in spatial navigation within an immersive virtual environment. Med. Sci. Monit. **15**(4), 140–150 (2009)
30. Moffat, S.D., Elkins, W., Resnick, S.M.: Age differences in the neural systems supporting human allocentric spatial navigation. Neurobiol. Aging **27**(7), 965–972 (2006)
31. Yuan, P., Daugherty, A.M., Raz, N.: Turning bias in virtual spatial navigation: age-related differences and neuroanatomical correlates. Biol. Psychol. **96**, 8–19 (2014)

32. Zahr, N.M., Rohlfing, T., Pfefferbaum, A., Sullivan, E.V.: Problem solving, working memory, and motor correlates of association and commissural fiber bundles in normal aging: a quantitative fiber tracking study. NeuroImage **44**(3), 1050–1062 (2009)
33. Jansson, E.S.V.: TSBK07 Computer Graphics Project Report for NQ Sokoban (2016). https://caffeineviking.net/share/papers/rnqsok.pdf
34. Mueller, S.T., Perelman, B.S., Veinott, E.S.: An optimization approach for measuring the divergence and correspondence between paths. Behav. Res. Methods **48**(1), 53–71 (2016)
35. Salas, E., Shuffler, M.L., Thayer, A.L., Bewell, W.L., Lazzara, E.H.: Understanding and improving teamwork in organizations: a scientifically based practical guide. Hum. Resour. Manag. **54**(4), 599–622 (2015)
36. Cooke, N.J., Gorman, J.C., Myers, C.W., Duran, J.L.: Interactive team cognition. Cogn. Sci. **37**(2), 255–285 (2013)
37. MacMillan, J., Entin, E.E., Serfaty, D.: Communication overhead: the hidden cost of team cognition. In: Salas, E., Fiore, S.M. (eds.) Team Cognition: Understanding the Factors that Drive Process and Performance, pp. 61–82. American Psychological Association, Washington, DC (2004)
38. Mercado, J.E., Rupp, M.A., Chen, J.Y.C., et al.: Intelligent agent transparency in human–agent teaming for Multi-UxV management. Hum. Factors **58**(3), 401–415 (2016)
39. Selkowitz, A.R., Larios, C.A., Lakhmani, S.G., Chen, J.Y.C.: Displaying information to support transparency for autonomous platforms. In: Savage-Knepshield, P., Chen, J.Y.C. (eds.) Advances in Human Factors in Robots and Unmanned Systems, pp. 161–173. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-41959-6_14
40. Chen, J.Y.C., Procci, K., Boyce, M., et al.: Situation awareness-based agent transparency (ARL-TR-6905). US Army Research Laboratory, Aberdeen Proving Ground, MD (2014)
41. Barber, D., Lackey, S., Reinerman-Jones, L., Hudwon, I.: Visual and tactile interfaces for bi-directional human robot communication. In: Proceedings for the International Society for Optics and Photonics, SPIE Defence, Security, and Sensing conference. SPIE, Baltimore (2013)
42. Lackey, S., Barber, D., Reinerman, L., Badler, N.I., Hudson, I.: Defining next-generation multi-modal communication in human robot interaction. In: Proceedings of the Human Factors and Ergonomics Society, vol. 55, no. 1, pp. 461–464 (2011)
43. Rau, P.P., Li, Y., Li, D.: Effects of communication style and culture on ability to accept recommendations from robots. Comput. Hum. Behav. **25**(2), 587–595 (2009)
44. Selkowitz, A.R., Lakhmani, S.G., Larios, C.N., Chen, J.Y.C.: Agent transparency and the autonomous squad member. In: Proceedings of the Human Factors and Ergonomics Society, vol. 60, no. 1, pp. 1319–1323. Sage, Los Angeles (2016)
45. Stowers, K., Kasdaglis, N., Rupp, M., Chen, J., Barber, D., Barnes, M.: Insights into human-agent teaming: intelligent agent transparency and uncertainty. In: Savage-Knepshield, P., Chen, J.Y.C. (eds.) Advances in Human Factors in Robots and Unmanned Systems, vol. 499, pp. 149–160. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-41959-6_13
46. Wright, J.L., Chen, J.Y.C., Barnes, M.J., Hancock, P.A.: Agent reasoning transparency: the influence of information level on automation-induced complacency (ARL-TR-8044). US Army Research Laboratory, Aberdeen Proving Ground, MD (2017)

Education, Training and Simulation

A Maximum Likelihood Method for Estimating Performance in a Rapid Serial Visual Presentation Target-Detection Task

Jonroy D. Canady[1(✉)], Amar R. Marathe[1], David H. Herman[2], and Benjamin T. Files[3]

[1] Human Research and Engineering Directorate, Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
{jonroy.d.canady.civ,amar.r.marathe.civ}@mail.mil

[2] PsyML, Los Angeles, CA 90017, USA
dh.herman@gmail.com

[3] Human Research and Engineering Directorate, Army Research Laboratory - West, Los Angeles, CA 90094, USA
benjamin.t.files.civ@mail.mil

Abstract. In human-agent teams, communications are frequently limited by how quickly the human component can deliver information to the computer-based agents. Treating the human as a sensor can help relax this limitation. As an instance of this, the rapid serial visual presentation target-detection paradigm provides a fast lane for human target-detection information; however, estimating target-detection performance can be challenging when the inter-stimulus interval is short, relative to human response time variability. This difficulty stems from the uncertainty in assigning each response to the correct stimulus image. We developed a maximum likelihood method to estimate the hit rate and false alarm rate that generally outperforms classic heuristic-based approaches and our previously developed regression-based method. Simulations show that this new method provides unbiased and accurate estimates of target-detection performance across a range of true hit rate and false alarm rate values. In light of the improved estimation of hit rates and false alarm rates, this maximum likelihood method would seem the best choice for estimating human target-detection performance.

Keywords: RSVP · Hit rate · False-alarm rate · Response time · Simulation

1 Introduction

Realizing an effective human-agent team (HAT) is a challenging problem for several reasons [1]. A common approach to bridging the wide gap in characteristics and capabilities between human and nonhuman agents is to force the computer agents to emulate human communication patterns, effectively limiting the data stream such that a human can easily parse it. In many contexts, operating solely at the speed of traditional human input and output is sufficient; however, some HATs could benefit greatly from a faster interface. Treating the human as a sensor is one way to speed up communication from human to agent.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 383–392, 2018.
https://doi.org/10.1007/978-3-319-91581-4_28

Whether reading real-time physiological signals or viewing human output with varying levels of confidence, modeling the human as a sensor allows computer agents to take action beyond responding to discrete commands [2]. By evaluating the past, present, and future state of this human sensor, many decisions and actions can be performed at the agent's pace, rather than the human's. One instance of this would be the application of the rapid serial visual presentation (RSVP) technique to the context of target detection.

Finding target images (e.g., images of threats) in a large collection of video and imagery data is a difficult problem. While computer vision algorithms can perform this task very rapidly and well in many environments, humans are currently far faster at adapting to significant contextual changes in the task (e.g., the landscape shifts from rural to urban, targets of interest change from fighter aircraft to insurgents carrying concealed weapons, etc.). A HAT can leverage the strengths of the human and the agent for target detection [3], and RSVP has been shown to enable humans to find target images much faster than self-paced viewing [4, 5].

In applying RSVP to a HAT's target detection task, the goal is to identify targets in the database of images the agent has labeled with significant uncertainty. However, some of the uncertainty can be reduced, if the factors that affect human performance in RSVP paradigms are known. Task performance can be characterized in experiments with known image labels. Unfortunately, it can be challenging to quantify such performance when response time variability exceeds the inter-stimulus interval [4, 6]. RSVP target detection performance is commonly characterized by hit rate (HR) and false alarm rate (FAR), but determining whether a given response is a hit or false alarm requires attributing the response to a particular stimulus, and response time variability makes such attribution difficult.

The need to determine whether a response is a quick reaction to the most recent stimulus or a slow reaction to the prior stimulus has given rise to a couple common attribution methods. The windowing method establishes a window of time after each target stimulus and determines any response in that window to be a hit. Alternatively, the distribution approach estimates a response time probability density function (RT-PDF) to assign responses to stimuli. Previously, we have described a regression method that builds on the distribution approach to provide better estimates of HR and FAR, particularly when the stimulus presentation rate is high and/or the FAR is non-zero [7]. Unfortunately, this regression method uses a heuristic based on the windowing method to estimate the RT-PDF, so it tends to lose performance under the same conditions that cause the windowing method to perform poorly.

Here, we introduce a maximum likelihood estimation (MLE) method that generally outperforms both the windowing method and the regression method, supplying more accurate estimates of the HR, FAR, and RT-PDF. These more accurate estimates of target detection performance can also improve the detection of target images in applications where it's not known in advance which images are targets or nontargets.

2 Methods

2.1 Estimation Methods

Windowing and distribution are two classes of methods commonly used to estimate HR and FAR in RSVP target detection tasks. Previously, we described a regression method to improve such estimates [7]. Now we describe an MLE method to further improve on HR and FAR estimation.

Established Methods

Windowing and Distribution. The windowing approach labels as a hit any response that falls within a certain window of time (typically from 0 to 1000 ms post-target) after a target stimulus. Any responses that don't match that criteria are labeled as false alarms. The number of targets hit is then divided by the total number of targets to compute the HR, while the FAR is determined by dividing the number of false alarms by the number of nontarget stimuli. Different implementations of this method vary in how responses are scored when more than one response falls within a response window and when a response falls within more than one response window.

The distribution approach makes use of an estimate of the response-time distribution to assign responses to specific stimuli [8]. The estimated value of a given stimulus's RT-PDF at the time of a response determines the likelihood that the response was caused by that stimulus. The likelihood for each potentially causal stimulus is divided by the sum of the likelihoods for all such stimuli for normalization [9]. Figure 1 depicts these two methods.

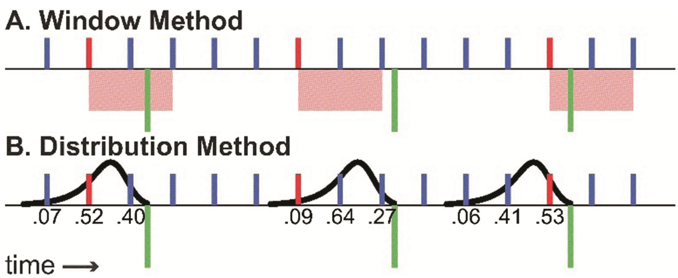


Fig. 1. Timelines illustrating existing response assignment methods. Blue, red, and green hash marks represent onset times of non-target stimuli, target stimuli, and responses, respectively. (Color figure online)

Regression Method. Our previously described regression method is based on the distribution method using an apportionment function. The distribution method produces biased estimates of HR and FAR, and the size and direction of those biases depend on the true HR and FAR. By regressing out those errors, we get better estimates. We have shown previously how the regression method outperforms the windowing and distribution methods, but it has two drawbacks. First, the regression method assumes normal error variance of the apportionment function, but that is not tenable, given that it is bounded between 0 and 1. Second, it uses a heuristic derived from the logic of the window method to estimate the response-time distribution, which can introduce inaccuracies from that method.

Proposed Method. Here we detail using an MLE to improve HR and FAR estimation beyond the regression method. In order to carry out MLE, a model is needed that provides the probability of getting a particular result given some parameter values. The results of an RSVP experiment are the set of times (B) at which a button was pressed. The probability that a button press occurs at a given time depends on the probability that a preceding stimulus elicits a response and the probability that a response to a particular stimulus would land at that time.

The probability that a stimulus (S_i) elicits a response is modeled as the hit rate (h) when the stimulus is a target (i.e., $s_i \in T$) and the false alarm (f) rate when the stimulus is a nontarget (i.e., $s_i \in NT$). The response time probability density function is modeled as an ex-Gaussian [10], which describes the probability density function of the sum of a Gaussian random variable and an exponential random variable. It therefore has three parameters: μ and σ describing the Gaussian variable and τ describing the exponential variable.

To summarize, the model is parameterized by h, f, μ, σ, and τ. Let $\theta = \{h, f, \mu, \sigma, \tau\}$. The probability that a given stimulus s_i elicits a response at time t parameterized by θ is:

$$P(s, t;\theta) = \begin{cases} 0, \delta t \leq 0 \\ h \times f(\delta t;\mu, \sigma, \tau), s_i \in T \\ f \times f(\delta t;\mu, \sigma, \tau), s_i \in NT \end{cases}, \quad (1)$$

where $f(\delta t; \mu, \sigma, \tau)$ is the ex-Gaussian density function, and δt is the difference between the onset of s_i and t. To compute the probability that one (or more) responses occurred at a time, we compute the complement of the probability that no response occurs at that time:

$$P_r(t;\theta) = 1 - \prod\nolimits_i 1 - P\left(s_i, t;\theta\right). \quad (2)$$

With this function, we can compute the probability of an overall result B as:

$$P(B;\theta) = \prod\nolimits_{t \in \mathrm{B}} P_r(t;\theta) \times \prod\nolimits_{t \notin \mathrm{B}} 1 - P_r(t;\theta). \quad (3)$$

To compute the second term in the preceding equation, we discretize time by dividing the entire time-course of the experiment into windows of time (time bins) spanning 10 ms each. We have found that increasing the time resolution beyond this value has limited practical effect and increases analysis time dramatically.

The hit rate and false alarm rate are related via the total number of responses:

$$h \times N_T + f \times N_{NT} = N_B. \quad (4)$$

The number of targets (N_T) and the number of non-targets (N_{NT}) are known. The number of button press responses in the result, N_B, can be considered fixed. As a consequence, only h need be estimated rather than both h and f. This simplifying assumption reduces the number of parameters in the model, reducing variance of the estimate.

Given this model of the probability of a result parameterized by h, f, μ, σ, τ and with a fixed N_B, a minimizer is used to find the values of the parameters θ that minimize $-\ln P(B;\theta)$. We used MATLAB's fminsearch, using a logistic linker function for h and log linker for μ, σ, and τ. Initial values for these parameters were drawn randomly from a beta function with parameters 2 and 0.65 for h and from lognormal distributions with mean and standard deviation of -1.5, 0.4; -1.5, 0.3; and -2.2, 1 for parameters μ, σ, and τ, respectively. Because fminsearch is not guaranteed to arrive at a global minimum, it should be run several times with different initial values to ensure convergence. In our simulations, we compared the log likelihood of the solution obtained from fminsearch to the log likelihood of the simulated parameters, and if the likelihood of the estimate was more than 1 log unit worse than the likelihood of the true parameters, the minimizer was re-run. A MATLAB implementation of this MLE method will be available at https://github.com/btfiles/RPE.

2.2 Evaluating the Methods

To evaluate the MLE method, we ran simulations to compare its performance against the traditional windowing method and our previous regression method. This consisted of simulating responses based on known HRs and FARs and then analyzing the resulting data with all three methods to determine the accuracies with which they recover the HR and FAR for various realistic true values. The 25 conditions simulated were comprised of five values of HR (0.50, 0.75, 0.90, 0.95, and 1.00) combined with five values of FAR (0.1, 0.05, 0.01, 0.005, and 0.001), and each condition was simulated 500 times to collect performance statistics. All simulations and analyses were performed using custom scripts in MATLAB versions 2015a and 2017a (MathWorks, Natick, MA).

Simulating the Responses. For each simulation, we generated a randomized RSVP experiment with a stimulation rate of 4 Hz, a target proportion of 10%, and a total stimulus count of 2400. Then, a random subset of all targets and non-targets was selected to generate responses such that the simulated rate of each block was as close as possible to the HR and FAR for the condition, while still having a whole number of responses. When a response was generated, a random draw was taken from an ex-Gaussian distribution (where $\mu = 0.35$, $\sigma = 0.10$, and $\tau = 0.25$), and a response event was added at that time after the generating stimulus.

Analyzing the Simulations. After simulating all of the responses necessary to generate the target HRs and FARs, the resulting stimulus and response timelines were analyzed using the windowing, regression, and MLE methods described above. Each simulation generated an estimate of HR and FAR from each method, and estimates of the response-time distribution parameters were generated by the regression and MLE methods. Each of these estimates was compared against the known simulated value to assess estimation error. Estimation error is summarized as root mean squared error (RMSE) and mean estimation error. The former provides an estimate of how far the estimator is likely to be from the true value, and the latter indicates whether the estimator is biased toward positive or negative errors.

3 Results

3.1 Performance Per Condition

For each simulation, the estimation errors for each method were calculated as the estimated rate minus the "true" simulated rate. The resulting 500 error values per rate type per condition per estimation method are depicted in Fig. 2 as box plots. Similarly, the error values for the regression and MLE methods' estimations of the three response-time distribution parameters are shown here.

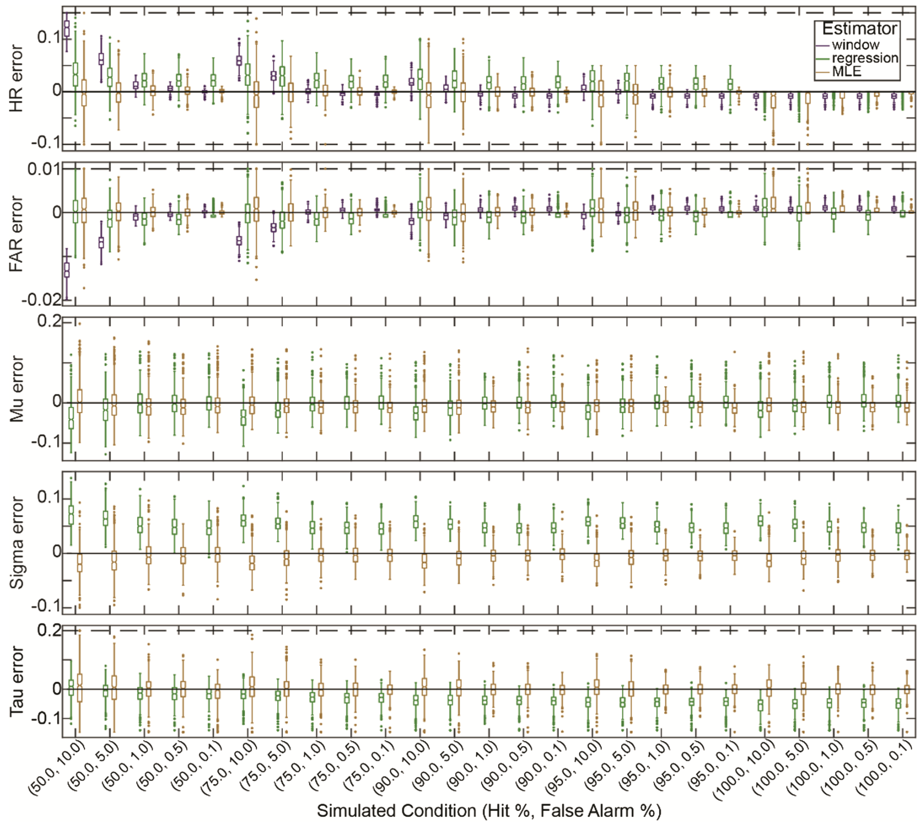


Fig. 2. Estimation errors of windowing, regression, and MLE methods for each condition.

For many conditions, the MLE method provides superior estimates of the HR and FAR, but it tends to have a wider range of variation in cases where the HR is 90% or higher and the FAR is 5% or higher.

3.2 Aggregate Performance

Averaging across all conditions, we obtain the aggregate performance shown in Fig. 3 and Table 1. While the windowing and regression methods tend to overestimate the HR (respective errors 8.771×10^{-3} and 1.6457×10^{-2}) and slightly underestimate the FAR (respective errors -9.0×10^{-4} and -5.8×10^{-4}), the MLE method estimates the HR and FAR with almost no bias either way (respective errors -2.52×10^{-3} and 3.58×10^{-4}).

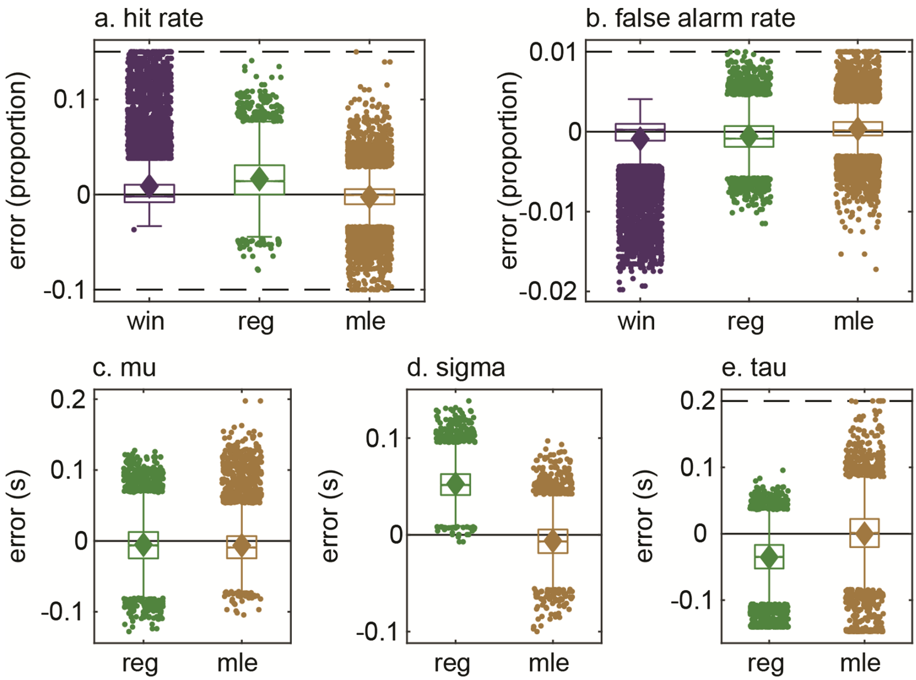


Fig. 3. Estimation error collapsed across conditions. The appearance of so many outliers is reasonable, given that each method's box plot is comprised of 12,500 estimations.

Table 1. Average estimation error collapsed across conditions. The response-time distribution parameters are not estimated by the window method.

	HR	FAR	Mu	Sigma	Tau
Win	0.008771	−0.0009			
Reg	0.016457	−0.00058	−0.00529	0.052683	−0.03526
MLE	−0.00252	0.000358	−0.00632	−0.00618	−0.00015

Table 2 gives RMS errors for the various estimates, showing that the MLE method's HR and FAR estimates are the most accurate, with respective RMS errors of 0.021508 and 0.002393, when compared to the windowing and regression methods' RMS errors for HR (0.032351 and 0.027795, respectively) and FAR (0.003542 and 0.00243, respectively). Also, the MLE method provides more accurate estimates of all three response-time distribution parameters than the regression method. While both methods provide similar estimates of mu, the regression method exhibits noticeably worse performance when estimating sigma or tau.

Table 2. RMSE of estimates collapsed across conditions. The response-time distribution parameters are not estimated by the window method.

	HR	FAR	Mu	Sigma	Tau
Win	0.032351	0.003542			
Reg	0.027795	0.00243	0.031954	0.055341	0.046857
MLE	0.021508	0.002393	0.030272	0.020898	0.040894

4 Discussion

The purpose of these simulations was to measure the MLE method's performance at recovering the true simulated HR and FAR, and the results above show that the MLE outperforms both the classic windowing method and the regression method. While all three methods can exhibit estimation error in excess of 10%, these inaccuracies are not equivalent.

The windowing method tends to overestimate the HR when the true HR is relatively low and/or the true FAR is relatively high. This is to be expected, thanks to its simplistic benefit-of-the-doubt approach. This approach is vulnerable to incorrectly classifying false alarms as hits when they fall too close to a target. On the other hand, this method performs very well when the human makes very few mistakes, such that the true FAR is very low.

While the regression method generally makes smaller estimation errors than the windowing method, it is still somewhat susceptible to overestimating the HR in the same situations that the windowing method does. This is because the regression method estimates the RT-PDF using a windowing-based heuristic, which partially counteracts the gains from using linear regression to correct such systematic errors.

The MLE method provides even more accurate HR and FAR estimates than the regression method, in part because it can simultaneously estimate the HR, FAR, and RT-PDF. Furthermore, the regression method is hampered by its treatment of non-Gaussian error as Gaussian. Overall, this results in the MLE method providing the most accurate performance estimates, which more strongly resist the effects of varying true values of HR and FAR.

4.1 Assumptions and Tradeoffs

The MLE method makes some assumptions based on the data model, including a constant hit rate, a fixed number of responses, and response times which are identically and independently distributed as an ex-Gaussian. However, we know that responses aren't strictly independent because simultaneous responses are treated as a single response and because the attentional blink phenomenon typically prevents the human from perceiving images that appear shortly after a target image [11, 12]. Furthermore, HR and FAR probably aren't strictly constant over the course of an experiment, since some targets are simply easier to detect than others.

In cases of very high HR and very low FAR, the regression method can appear to outperform the MLE method. This is likely an effect of the MLE representing proportions as the logistic transformation of real numbers, which maps positive and negative infinity to zero and one, respectively. This means that a minimizer can achieve results that are arbitrarily close to zero and one, but it will never equal those values. In contrast, the regression approach simply truncates estimates to keep them between 0 and 1. This truncation can make the regression method appear more accurate.

When the true HR is very high, the true FAR is very low, and responses are mostly faster than 1 s, the human's performance upholds the heuristic of the windowing method very well. This enables the windowing method to provide highly accurate estimates of HR and FAR that outperform the MLE method. However, in many contexts, such high human performance is unlikely, and it would indicate that the human's throughput was not even approaching capacity. Thus, based on its overall better estimation of HR and FAR, the MLE method proposed here would seem the best choice for estimating HR and FAR

4.2 Application

In real-world applications, such as improving a HAT tasked with target detection, the goal of the RSVP target detection paradigm is likely to identify target images from a large set of imagery. Not knowing the target status of the images precludes direct application of the MLE method. However, previous solutions to this problem, such as a Bayesian formulation to estimate target probability [8], would benefit from the more accurate HR and FAR estimates provided by the proposed MLE method. Furthermore, the more accurate estimate of the response time distribution derived from the MLE method should improve the performance of image classification based on button presses using methods derived from the distribution method.

References

1. Klien, G., et al.: Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intell. Syst. **19**(6), 91–95 (2004)
2. Wang, D., et al.: Using humans as sensors: an estimation-theoretic perspective. In: Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, pp. 35–46 (2014)

3. Bohannon, A.W., Waytowich, N.R., Lawhern, V.J., Sadler, B.M., Lance, B.J.: Collaborative image triage with humans and computer vision. In: IEEE Systems Man and Cybernetics, pp. 004046–004051 (2016)
4. Mathan, S., Ververs, P., Dorneich, M., Whitlow, S., Carciofini, J., Erdogmus, D., et al.: Neurotechnology for image analysis: searching for needles in haystacks efficiently. In: Augmented Cognition: Past, Present and Future (2006)
5. Parra, L.C., Christoforou, C., Gerson, A.D., Dyrholm, M., Luo, A., Wagner, M., et al.: Spatiotemporal linear decoding of brain state. IEEE Signal Process. Mag. **25**(1), 107–115 (2008)
6. Sajda, P., Pohlmeyer, E., Wang, J., Parra, L.C., Christoforou, C., Dmochowski, J., et al.: In a blink of an eye and a switch of a transistor: cortically coupled computer vision. Proc. IEEE **98**(3), 462–478 (2010)
7. Files, B.T., Marathe, A.R.: A regression method for estimating performance in a rapid serial visual presentation target detection task. J. Neurosci. Methods **258**(30), 114–123 (2016)
8. Gerson, A.D., Parra, L.C., Sajda, P.: Cortically coupled computer vision for rapid image search. IEEE Trans. Neural Syst. Rehabil. Eng. **14**(2), 174–179 (2006)
9. Marathe, A.R., Lance, B.J., Nothwang, W., Metcalfe, J.S., McDowell, K.: Confidence metrics improve human-autonomy integration. In: Proceedings of 9th ACM/IEEE International Conference on Human-Robot Interaction (2014a)
10. Lacouture, Y., Cousineau, D.: How to use MATLAB to fit the ex-Gaussian and other probability functions to a distribution of response times. Tutor. Quant. Methods Psychol. **4**(1), 35–45 (2008)
11. Raymond, J.E., Shapiro, K.L., Arnell, K.M.: Temporary suppression of visual processing in an RSVP task: an attentional blink? J. Exp. Psychol. Hum. Percept. Perform. **18**(3), 849–860 (1992)
12. Shapiro, K.L., Raymond, J.E., Arnell, K.M.: The attentional blink. Trends Cogn. Sci. **1**(8), 291–296 (1997)

Virtual Reality Training to Enhance Motor Skills

Matt Dombrowski(✉), Ryan Buyssens, and Peter A. Smith

University of Central Florida, Orlando, USA
{MattD,Ryan.Buyssens,Peter.smith}@ucf.edu

Abstract. The use of Virtual Reality (VR) and Augmented Reality (AR) as a healing aid is a relatively newer concept in the field of rehabilitation and training. Clinicians now have access to virtual worlds and games in which they can immerse their patients into interactive scenarios that would not have been possible in previous years.

Studies have shown effective results when VR/AR is incorporated into rehabilitation and training therapy practice. Mobility limited individuals can move freely within an open world virtual environment using the enhancements of virtual reality platforms like the HTC VIVE and Oculus Rift. The recent widespread consumer availability of VR/AR platforms has made it possible for clinicians to have the ability to incorporate fully immersive tech into their treatment regimens. Their research methods range from the implementation of consumer video game systems to custom developed hardware and software to enhance training. Clinicians are utilizing VR/AR platforms to better engage their patients. In doing so they are improving the effectiveness of training. Researchers have seen the implementation of these new tools improve the psychological effects of phantom limb syndrome, and improve motor skills for those with multiple sclerosis, cerebral palsy and other mobility debilitating conditions.

This research will survey the past, present and future of applications and research in VR/AR Game experiences to aid in training and rehabilitation, exploring current state of research and the documented effectiveness of using games to heal. The benefits and potential further uses for emerging technologies within the healthcare field will guide the implementation of VR/AR applications to aid in the training of children to best learn to implement 3D printed prosthetic limbs in their everyday lives. In collaboration with Limbitless Solutions at the University of Central Florida, researchers have been engaged in discussions with young recipients of 3D prosthetic limbs. The researchers will discuss their findings and plans for how VR/AR Game experiences can provide solutions for the enhancement of the day to day routines of young prosthetic users.

Keywords: Games · Virtual Reality · Augmented Reality · Game design
Motor skills

1 Introduction

Virtual Reality (VR) has been in the laboratory for years, the realistic experiences that these computer simulations have provided to researches was incredibly exciting in the 1980s, but due to low resolutions, and large screens it was difficult to bring virtual reality

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 393–402, 2018.
https://doi.org/10.1007/978-3-319-91581-4_29

out of the laboratory and into the hands of consumers. Recent advances in virtual reality, powered mainly by the advent of small, high resolution screens, with high enough refresh rates to make VR glasses light enough to be comfortable to the average consumer. Additionally, the cost has come down to below $1000.00 plus the cost of a new PC. Surprisingly, while VR has become consumer friendly, Augmented Reality (AR) has in many ways surpassed it, utilizing similar technology but with see through, or video backed lenses.

As these technologies have matured their use in consumers lives, as well as in the research laboratory is growing. Especially given the new-found likelihood of applications being adopted by users. This exciting opportunity has led to the design of VR and AR experiences to potentially help train limb deficient children to use their prosthetic arms, as well as provide an opportunity for others to experience what like is like for them today.

2 Background

The history of immersive technologies is a long and storied one. Though the actual term "Virtual Reality" was termed by Jaron Lanier in 1987 [1], the idea behind immersive experience dates back over 100 years. As early back as the nineteenth century painters were painting 360-degree panoramic art pieces [2].

The late nineteenth century also brought the advent of stereoscopic technology in the form of viewers and toys. This would lead to the popular 20th century toy the View-Master, a stereoscope (patented 1939), was used for "virtual tourism" [3].

Even the Link Trainer, designed in the 1930's first found popularity as a carnival ride before being accepted as the first use of immersive technology incorporated into training. As the first foray in flight simulators, the "Link Trainer", developed by Edward Link, was a fully electromechanical simulator that would eventually be used to train pilots for World War II. At first users were skeptical, which left the device only really useful as a toy, but during the war pilots were having problems landing due to heavy fog on the eastern front, the need to train for instrument only landing became apparent and the Link Trainers were a huge success for training pilots to land [4] (Figs. 1 and 2).

Fig. 1. Sevastopol Panorama (created in 1905), depicting the Siege of Sevastopol [2]

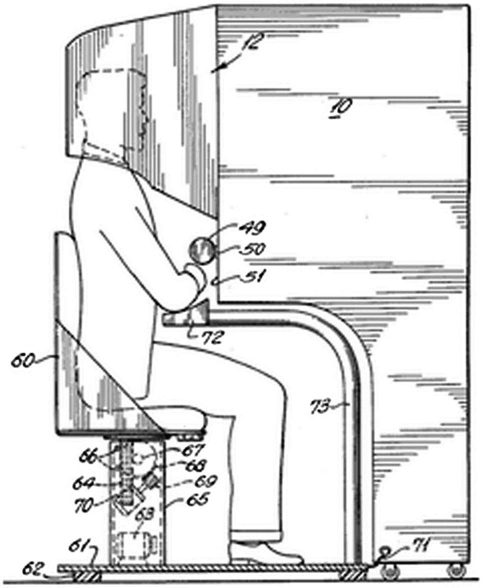


Fig. 2. Morton Heilig's Sensorama [5]

During the 1950's, Morton Heilig's Sensorama that provided the illusion of reality with smell, stereo sound, simulated real world effects like wind and vibration for up to four people at a time. Heilig also developed the first head mounted displays Heilig's development and research has earned him the moniker the "father of virtual reality" [5].

The mid to late 1960's Ivan Sutherland created a concept that would drive the future of the VR industry, known as the "Ultimate Display." The idea behind the ultimate display is a room in which a computer could control matter, if it exists, where, and how. In such a space a virtual object like a chair could be sat on, chains could imprison, bullets could be fatal. At the time such a room was described as similar Wonderland from the Alice in Wonderland Books. The most obvious comparison today would be the Holodeck from Star Trek [6].

It was not until the late 1960s that Myron Kruegere developed interactive art with VR in which he termed "Artificial Reality." [7] It was not until around 20 years later during the late 1980's that a term would be coined to encompass all the past and future research in this field. Jaron Lanier, founder of the visual programming lab not only coined the term "Virtual Reality" but also played a major role in the development of haptic feedback [1].

From this point to the early 2000's Virtual and Augmented reality truly had its revolution. VR and AR products were being developed, not only for simulation, but also for entertainment value. The 1990's saw commercial VR developments from video game powerhouses SEGA and Nintendo. It also brought substantial price drops in the hardware factor of these devices. Both Nintendo and Sega provided products, though flawed, under the $200 price point [8].

The Nintendo Virtual Boy, was Nintendo's Virtual Reality follow up to the Amazingly popular Game Boy. Designed by Gunpei Yokoi, the designer of most of Nintendo's major hardware releases from the Game & Watch series of mobile games until the ill-fated Virtual Boy. Gunpei Yokoi's design philosophy was to use withered technology in new ways, that is to say he would find tested and market ready technology. The Virtual Boy, did not follow this philosophy. It used a red and black only display that was not ready for the consumer market. The screen gave users headaches, and the Virtual Boy is known as one of Nintendo's first major failures in the hardware arena. It also led to Gunpei Yokoi leaving the company shortly after [8].

The rise in augmented reality (AR) has really stemmed from the recent mass availability of AR ready technology. From AR Smartphones, Automotive HUD Displays, and the next generation of AR Headsets, users are being introduced to AR in their everyday routines. AR was first introduced via military applications in the early 1990's [9].

This brings us to today's VR and AR market, "Global virtual reality revenues will reach $7.17 billion by the end of this year, according to a new report by Greenlight Insights". They also estimate that the industry will rise to over 75 billion dollars in 2021. Companies such as Samsung, Google, Apple Oculus and HTC have all put their hats into the VR and AR market. The advent of mobile technology has allowed everyday consumers to have easier access to VR/AR capable hardware to utilize the tech in their own home [10].

3 The Current State of the Art in Training and Rehabilitation

VR and AR have a long tradition of being used in both training and rehabilitation. The current state of the art in these fields is laying the ground work for future applications to support prosthetics training.

3.1 Augmented Reality

AR technology offers many advantages in both training and learning environments [11]. AR's allows users to interact, collaborate and communicate in hybrid, digital and physical spaces. Current trends in tech actual state that augmented reality is set to take over the virtual reality markets in years to come. Mark Zuckerberg, founder of FB states, "The phone is probably going to be the mainstream consumer platform [where] a lot of these AR features become mainstream, rather than a glasses form factor that people will wear on their face." [12] This downsizing of equipment hardware allows AR to be instituted into more non-commercial fields such as research and rehabilitation.

AR is commonly used in the support of maintenance in industrial applications. This commonly is implemented as a world overlay generated by looking through a see translucent screen [13]. The technology is becoming smaller and less expensive. Microsoft entered this market with the Hololens which has been used in many applications including similar assembly applications, data visualization, and entertainment as well [14]. AR hardware has been found to be used in laparoscopic surgical training, neurosurgical procedure training, and echocardiography, however, there is less data available concerning transfer of learning, despite promising results existing [15].

AR technology is currently being researched in the medical field in conjunction with the treatment of multiple conditions, such as, extremity pain, phantom limb pain, and post stroke motor-rehabilitation. A recent study conducted by Brazilian physiotherapists, an augmented reality environment was utilized where participants could see themselves and their surroundings, as in mirror. They were asked to perform a series of tasks with and without the VR-mirror. The survey scores suggested greater motor improvement in the augmented reality group than in the control. These studies are intriguing from a training and education standpoint due the rising availability and lowering price points of AR tech. This should allow participants to have faster access to AR in home, as well as in hospital environments [16].

AR has been applied to post stroke finger rehabilitation. An AR training environment was used in conjunction with an assistive device to train finger extension. After 6 weeks of training there was a trend towards the reduction of assistive force in the impaired hand. There is encouraging evidence to continue this type of training [17].

3.2 Virtual Reality

VR is also being used in a variety of training and rehabilitation applications across military, education, healthcare, and industry. Virtual Reality has experienced a recent boom in the consumer level that has resulted in a resurgence of interest in the research lab as well.

Researchers at Aalborg University found that Virtual Reality could be used to reduce phantom limb pain. This happens when the VR environment tricks the amputee's brain into thinking that it is still able to control the missing limb. Phantom limb pain is thought to occur with the brain no longer receives feedback from a part of the body it expects to. The VR environment made the participant brain think it was still receiving feedback even when it wasn't [18].

In the treatment of phantom limb pain, a condition where amputees feel body parts that no longer exist, a team at the University of Tokyo Hospital has been using VR with much success. Masahiko Sumitani at University of Tokyo Hospital, has succeeded in using VR to treat phantom limb pain, a condition in which amputees or people with damaged nerves still feel pain from body parts that no longer exist. "We've found a mechanism through which pain can be reduced when patients are able to get an image of moving (their absent limbs) in their minds. This is something new," Sumitani said. VR technology allows amputees to see and manipulate virtual images of their lost limbs. The virtual limbs move accordingly with the intact parts of the arm, allowing the images to take root in patients' minds [19].

VR has also been used for stroke rehabilitation. Patients who obtained VR rehabilitation had a significant increase in upper limb motor impairments and motor relation functional abilities then patients that were rehabilitated using traditional methods [20].

Embodied Lab is an immersive project stemmed from a daughter's firsthand experience taking care of her ailing mother who was suffering from early-onset Alzheimer's disease. The mission of this project is to allow health care providers step into the perspectives of the patient and other members of the patients care team. This same patient-centric approach could be ported to those with missing limbs. Understanding the daily task of a limb deficient person would allow clinicians, developers and caregivers to better create accessible platforms and/or adapt to an accessible lifestyle [21].

Researchers in Seoul found VR to be useful to provide training for repetitive tasks and developed a mobile game-based virtual reality rehabilitation program for upper limb dysfunction after ischemic stroke. They found improvement from the VR intervention, but the results were not statistically significant. This is different from the AR similar AR research in that the difference was not significant. This is most likely a product of VR recreating reality while AR enhances it [22].

VR has also been found useful in rehabilitation in Parkinson's disease, specifically on tasks that could improve balance in users. Over a 12 week intervention participants using Wii remote based training reported more confidence in their ability to balance.

3.3 Current Uses of VR/AR in Gaming

As previously mentioned, the Microsoft Hololense is bringing the AR experience to the masses. This product has fueled the low-cost AR market by providing an attainable solution. Unfortunately, the Hololense suffers from an anemic field of view. This makes the most interesting applications out of reach. There is still solid support for the Unity game engine and the ability to play games. The perceived value of these applications is much higher given the mixed reality demonstrations that show a person interacting with Minecraft, a popular video game, while it covers their coffee table. In reality the field of view would not allow the user to view the whole coffee table at the same time.

A more common use of AR in games is through the use of AR cards. These are markers that a camera on a device like a phone, or Nintendo 3DS can pick up on their camera and use to add three-dimensional imagery to the scene. AR cards came with both the 3DS and PlayStation Vita. They also power the card game version of Skylanders for cell phones. One technology that powers this type of interaction Vufloria has evolved to use actual toys and other objects as AR markers.

The AR overlay glasses have become less used, as Google Glass has ended production. There are however competitors entering the marketplace, and the company Magic Leap, who received hundreds of millions in investment to build the next generation of AR for consumers has yet to release a product. AR is on the cusp of becoming the next big thing, but it is still hiding behind the major and more visible strides of VR.

What is becoming even more popular in AR is the use of AR stickers or graphics being added to pictures and photos. Popular services like SnapChat and Facebook are allowing users to change their faces to look dressed up like popular movies, of funny stretch and squished faces. These lenses are changing the way users share moments with each other, by being able to be silly or physically manifest emotion.

There are many commercially available systems for VR ranging from professional-level which can only be attained with large budget expenditures to the consumer level where some products can be had for as little as 20 dollars. On the stratospheric end of the professional level, providing an experience that is fully immersive consisting of a room of 360-degree projections and use 3D glasses to emphasize the virtual environment.

Another high-end experience is the VOID which includes VR headsets and haptic wearables worn in an environment that is mapped within the VR so that doors and hallways are replicated within the experience. This experience is location based and utilizes hardware that is far beyond what consumers can have at home, but the differentiation may be reduced as the next generation of consumer technology is released.

Already the consumer level are headsets like the PlayStation VR, Oculus Rift and HTC VIVE that are extremely affordable and can give a relatively immersive VR experience. The HTC VIVE's room scale VR is the closest thing consumers have to the embodied experience provided by the VOID. The other consumer experiences use front facing hand tracking making it difficult to turn around in the environment. The next generation will utilize inside out tracking, allowing the glasses to track the hands reducing this issue.

There are also attachments for cell phones like the Google Cardboard that, when coupled with apps, can convert the cell phone into a headset for VR use. The experiences provided by these devices are not nearly as full featured, mostly due to the quality of a phone compared to a $2,000.00 PC, but the line is blurring. The next generation of VR technology will be even more portable and affordable than ever before.

4 Design of a VR Environment for Prosthetic Training

Here are the University of Central Florida our research team is using AR and VR in various ways to create interactive training for underserved populations. This research utilizes gamification, custom video game controllers, AR and VR tech to aid in disability training and rehabilitation. This multi-modal approach allows our research team to develop broad and engaging interactions to keep user interest but still accomplish training. We are working with Limbitless Solutions Inc, a company that provides 3D printed prosthetics for children for no cost. As mentioned above the team developed a multitude of training tools to aid in the training associated with the prosthetics.

The team is incorporating AR stickers in the form of an interactive application that allows the user to select their customized prosthetic arm. This app allows children to customize and visually "test drive" their prosthetic before they receive it. Using mobile based tracking, researchers can track the virtual prosthetic on the video of the child utilizing a cell phone camera. This is a custom solution but works similarly to a SnapChat filter.

Our VR project comes from two vantage points that of the prosthetic users and users with two organic limbs. Our team is working on a VR game experience that will convey the concept of limb loss to those who have use of all of their limbs. The experience will immerse the player into a world where they may have an arm missing or no arms at all. They will have to perform tasks without the use of a limb and learn how to overcome the difficulty within the virtual game environment.

Concerning the prosthetic users, the VR training simulates exercises that fine tune aspects of timing and motor skills such that of a catch and release situation while tossing a ball. Utilizing custom created EMG video game controllers and trackers, researchers use the HTC VIVE to create a virtual space in which the user can use auditory and visual cues in efforts to improve their hand eye coordination. Our research goal is in conjunction with this training, prosthetic users will be able to more quickly improve their hand to eye coordination using their artificial limb.

Finally, a benefit in VR is role immersion. Developing for empathy and education is another way our research team is using VR. The team has created an interactive experience in which the player will only have access to one fully developed arm. The user has to perform a series of daily tasks that are timed. After the challenge the user is given full access both their limbs in efforts for them to related to the challenges limb deficient children face. The team is planning to provide, The Empathy Quotient (EQ), a pre and post assessment to determine a change in users empathy.

5 Conclusions

In conclusion, the authors have determined that VR and AR technology are current methods to enhance training for motor skill development. The researchers suggest a multi-modal approach when developing for motor skill training. The technology should drive the research never hinder it. The mixed mode approach in VR/AR is an ideal supplement to the current traditional methods of rehabilitation training.

Combining a solid foundation of research and prior art with new advances and commercial off the shelf technologies is allowing for better tools to train and rehabilitate users of prosthetic arms. The ability to bring prosthetics into the virtual world and allow users to train how to use there arm, while at the same time providing opportunities for empathy through having others experience what having a prosthetic arm would be like, is changing the way the world will think about prosthetics. It is clear that as VR and AR applications are built for prosthetic users new avenues will open to be explored. The future for this technology is bright.

5.1 Future Work

Another challenge arising in the near future is how to train for more advanced prosthetic gestures. The current iteration of the Limbitless arm simply opens and closes. The Limbitless team is in development of gestural controls and finger movements for their next release. This has posed a challenge for our games research team: How do you train a child to learn multiple, pattern based, electromyographic (EMG) inputs? Our team has developed a proprietary calibration system to simulate an analog input from a single EMG sensor. In addition, by using virtual reality and a 3D virtual limb, we are able to test feedback and EMG controlled gestures and simulate them in a virtual space. This allows our team to test various methods of input and output from the EMG mechanics. All doing so before porting the method of choice to the children user population.

References

1. VR Society: Who Coined the Term Virtual Reality?. https://www.vrs.org.uk/virtual-reality/who-coined-the-term.html Accessed on Feb 2018
2. Kaushic: Historic Panoramic Paintings and Cycloramas, Amusing Planet, 2061. www.bing.com/cr?IG=0BB14A839ABA4A8EB64E333867A8639F&CID=29211B57A04D6F13080010C9A1E26ED4&rd=1&h=WkD1nw-chyx0Qv8QiMi7GrvKp7PfIGzy9okj3NHnMiQ&v=1&r=http%3a%2f%2fwww.amusingplanet.com%2f2016%2f07%2fhistoric- panoramic-paintings-and.html&p=DevEx,5068.1
3. VR Society: History of Virtual Reality Virtual Reality Society (2018). www.vrs.org.uk/virtual-reality/history.html
4. De Angelo, J.: The Link Flight Trainer: A Historic Mechanical Engineering Landmark, Roberson Museum and Science Center, Binghamton, New York. http://www.mortonheilig.com/InventorVR.html Accessed on 10 June 2000
5. Sutherland, I.E.: The ultimate display. In: Multimedia: From Wagner to Virtual Reality, pp. 506–508 (1965)
6. Steinicke, F.: Being Really Virtual: Immersive Natives and the Future of Virtual Reality. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-43078-2
7. Stanton, R.: A Brief History of Video Games: From Atari to Xbox One. Robinson, London (2015)
8. Caudell, T.P., Mizell, D.W.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. 2, pp. 659–669. IEEE, January 1992. http://variety.com/2017/digital/news/virtual-reality-industry-revenue-2017-1202027920/
9. Cheng, K.H., Tsai, C.C.: Affordances of augmented reality in science learning: suggestions for future research. J. Sci. Educ. Technol. **22**(4), 449–462 (2013). https://doi.org/10.1007/s10956-012-9405-9
10. Merel, T.: The Reality of VR/AR growth, Techcrunch.com (2017). https://techcrunch.com/2017/01/11/the-reality-of-vrar-growth/
11. Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveriy, M., Bockholt, U., Tecchia, F.: Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interact. Learn. Environ. **23**(6), 778–798 (2015)
12. Evans, G., Miller, J., Pena, M. I., MacAllister, A., Winer, E.: Evaluating the Microsoft HoloLens through an augmented reality assembly application. In: Degraded Environments: Sensing, Processing, and Display, vol. 10197, p. 101970V. International Society for Optics and Photonics, May 2017
13. Barsom, E.Z., Graafland, M., Schijven, M.P.: Systematic review on the effectiveness of augmented reality applications in medical training. Surg. Endosc. **30**(10), 4174–4183 (2016)
14. Corrêa, A.G.D., Ficheman, I.K., de Deus Lopes, R., Klein, A.N., Nakazune, S.J.: Augmented reality in occupacional therapy. In: 2013 8th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1–6. IEEE, June 2013
15. Luo, X., Kenyon, R.V., Kline, T., Waldinger, H.C., Kamper, D.G.: An augmented reality training environment for post-stroke finger extension rehabilitation. In: 9th International Conference on Rehabilitation Robotics, 2005. ICORR 2005, pp. 329–332. IEEE (2005)
16. Aalborg University. Virtual reality eases phantom limb pain. ScienceDaily, 31 May 2017. www.sciencedaily.com/releases/2017/05/170531102921.htm Accessed on 21 Feb 2018
17. https://www.japantimes.co.jp/news/2016/10/09/business/tech/virtual-reality-helps-treat-phantom-pain-letting-missing-injured-limbs-move/#.Wo5Z-KinHZQ

18. Turolla, A., Dam, M., Ventura, L., Tonin, P., Agostini, M., Zucconi, C., Kiper, P., Cagnin, A., Piron, L.: Virtual reality for the rehabilitation of the upper limb motor function after stroke: a prospective controlled trial. J. Neuroeng. Rehabil. **10**(1), 85 (2013)
19. Embodied Labs: Embodied Lab About Page (2018). http://www.embodiedlabs.com/about/
20. Choi, Y.H., Ku, J., Lim, H., Kim, Y.H., Paik, N.J.: Mobile game-based virtual reality rehabilitation program for upper limb dysfunction after ischemic stroke. Restor. Neurol. Neurosci. **34**(3), 455–463 (2016)
21. Holmes, J.D., Gu, M.L., Johnson, A.M., Jenkins, M.E.: The effects of a home-based virtual reality rehabilitation program on balance among individuals with Parkinson's disease. Phys. Occup. Ther. Geriatr. **31**(3), 241–253 (2013)

Examination of Effectiveness of a Performed Procedural Task Using Low-Cost Peripheral Devices in VR

Damian Grajewski, Paweł Buń(✉), and Filip Górski

Chair of Management and Production Engineering, Poznań University of Technology, Piotrowo 3 STR, 60-965 Poznań, Poland
{damian.grajewski,pawel.bun,filip.gorski}@put.poznan.pl

Abstract. The paper presents a Virtual Reality (VR) training system dedicated for interactive course focused on acquisition of competences in the field of manual procedural tasks. It was developed as a response for the growing market demand for low-cost VR systems supporting industrial training. A scenario for the implementation of an elementary manual operation (modified peg-in-hole task) was developed. The aim of the test was to show whether the prepared solution (along with peripheral devices) can be an effective tool for training the activities performed at the production site. The procedural task was performed by specific test groups using various peripheral devices. The paper presents preliminary results of tests regarding evaluation of effectiveness of virtual training, depending on specific peripheral devices used.

Keywords: Virtual reality training · Interaction devices · Haptic feedback

1 Introduction

Virtual Reality (VR) is widely used in engineering education [1, 2], designing new complex mechanical systems [3], medical training applications [4, 5], prototyping of ergonomic workplaces [6] and advanced training and simulation systems [7, 8], among other things. Development of low-cost Virtual Reality interaction devices has been significant in recent years, which makes availability of hardware much higher than several years ago. These new devices (e.g. systems for tracking and gesture recognition or haptic devices [9]) allow users to directly interact with elements of a virtual scene in a more natural way and obtain a deep feeling of presence in an artificial environment. It is known as the immersion [10] and is one of the conditions that should be fulfilled when using Virtual Reality as a training tool.

In recent years, software previously used to create games was adopted to enable preparation of Virtual Environments. Popular game engines such as Unreal Engine or Unity 3D offer sets of plugins enabling integration of low-cost (consumer) VR hardware in any application. It makes it easier to create VR based games and training applications, but it also poses many new challenges.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 403–415, 2018.
https://doi.org/10.1007/978-3-319-91581-4_30

This paper discusses examples of using low-cost devices to build immersive training systems and compares effectiveness of training, depending on specific peripheral devices. Two different hardware systems were built, both based on low-cost VR projection and interaction devices, one of them using standard components (the non-haptic approach) and the other – a custom made haptic device in form of a Delta robot. The same educational application was tested on both hardware setups, on a group of users and the results of this testing are presented in the final part of the paper.

2 Materials and Methods

2.1 Learning Transfer

Transfer of learning or learning transfer is the amount of knowledge gained during training that can be applied to a new task [11]. Learning transfer can occur only when there is task similarity or domain knowledge and a given person is able to perceive the similarity [12]. Currently available VR systems offer possibility to create a real-like environment and/or situation. Based on the assumption, that skills acquired in a virtual environment can be transferred to a real situation, virtual reality can offer many training benefits:

- it enables simulation of difficult or dangerous work conditions or failures that cannot be reproduced with real equipment, thus enabling learning skills in safe conditions and avoiding downtime or damage of equipment [13];
- it allows making mistakes while learning procedural tasks, without any real consequences, what is particularly important in high-risk situations, such as airplane flying [14];
- it enables avoiding downtime of machines [15].

2.2 Virtual Reality in Industrial Training Applications

In industry, training of operators to use complex technical systems is an important source of expenditure. VR systems can be effective tools used in industrial training, reducing costs significantly. Analyzing the potential of this type of solutions, with particular regard to tactile interaction that occurs between the user and elements of the virtual environment, one can observe two approaches in its implementation. The immersive approach assumes use of stereoscopic visualization systems, which ensure a high level of immersion when used together with tracking systems. The downside of this solution is lack of a haptic feedback. In the immersive approach, devices for Rapid Prototyping (mostly via 3D printing) are increasingly used, thanks to which physical representations of digital models that are necessary during virtual simulation (e.g. models of hand tools, control panels) can be prepared [16, 17]. This increases the degree of immersion, but lack of the tactile stimulus limits the interaction with the user. The expected degree of tactile interaction can be provided by a haptic approach, which assumes the use of a manipulator with force feedback effect. In this approach, RP solutions are also used [18].

As a result of the analysis of existing approaches, it was noticed that the level of effectiveness of the interactive training simulation is determined by immersion (which can be achieved by capability of manipulation of virtual objects interactively) and tactile interaction. It is worth mentioning, that commercial solutions integrate immersive and haptic VR equipment, while allowing the manipulation of physical models (representations of digital data) to a small extent [19–23].

In the context of the above-mentioned disadvantages and limitations of existing solutions, as well as on the basis of the authors' own experience resulting from the implementation of research work [9, 18], it has become justified to develop similar but alternative to the haptic approach and comparing it with a solution based on non-contact interaction techniques.

2.3 Low-Cost VR Devices

A set of low-cost projection and interaction devices was used to build a demonstration station designed for tactile interaction studies. Proposed solution is based on a consumer HMD - Oculus Rift CV1 (parameters presented in Table 1). The HMD itself was used in both hardware setups. Also, a pair of dedicated controllers (Oculus Touch) were used for the non – haptic approach.

Table 1. Parameters of Oculus Rift CV1

Parameter name	Value
Resolution	2160 × 1200
FOV (Field of View)	110
Mass	470 g
Communication	USB for tracking and control, display by HDMI
Interaction	Controller - Oculus Touch
Tracking	IR LED sensor – 3 DOF positional tracking, 1,5 × 1,5 [m] tracking space, built-in accelerometer for 3 DOF rotational tracking
Price	500 USD

The optical PST-55 tracking system was used in the second hardware setup (the haptic approach), for interaction testing. It is a set consisting of a device for emission and detecting reflected infrared light, a camera registering visible light, and markers placed on a tracked object. The most important parameters of device are presented in Table 2.

The solution is based on passive markers, which are used determine the position and tracking of objects in the distance of about 0.4–3.2 [m] from the device [24]. For each object to be recognized by the system, an appropriate, unique, constant marker arrangement must be recorded into the tracking system software, using an appropriate procedure. The PST system enables to track linear movements in all three axes and object rotation relative to each axis.

Table 2. Parameters of PST-55

Parameter name	Value
Tracking accuracy (position)	<1 [mm]
Tracking accuracy (orientation)	<1 [degree]
Sensor type	IR Led sensor
Data Interface	VRPN, through SDK, export of data to csv format
Latency	18 [ms]
Number of tracked objects	Up to 15
Price	about 3000 USD

2.4 Low-Cost Haptic Device

The authors have prepared a solution based on the so-called an active touch device, whose role is played by a manipulator with a parallel kinematic structure (Delta type robot). The main task of the robot, as an alternative to the currently used haptics, is to simulate the shape of digital objects, and thus provide user with a tactile stimulus during the interaction [9]. Introducing a new element to a virtual environment - an active touch device – is an approach that has not been used in the creation of VR training systems so far. In comparison to existing solutions, use of the robot also provides considerable freedom of movement for the user (available systems are based on a permanent contact of user's hand with the working tip of the haptic manipulator, which has a small working range). Design of the device is based on kinematics of a stationary industrial robot of the so-called parallel structure 3 (RPR) [25] is crucial in a proposed system designed for touch interaction research in Virtual Reality. The basic design assumptions (including number of degrees of freedom, ranges of motion, forces, speed and positioning accuracy of the end effector) of the robot were developed as part of the research project VISIONAIR [26]. It was assumed that the working area of the device will be about $600 \times 600 \times 600$ [mm], the maximum speed in individual axes will not be higher than 100 [mm/s], and the maximum generated force will not exceed 10 [N].

On the basis of the mentioned assumptions, the drives, measuring and controlling elements of the manipulator were selected. Synchronous motors with permanent magnets PMSM were used as drivers. They were connected to the gear and coupled to a resolver. The robot is shown in Fig. 1.

Due to the fact, that the device was to carry out tasks in direct contact with the user, during its design, a special emphasis was put on the security issues (including flexible arms). The factors that decided about choosing mentioned device for the developed concept of the tactile interaction testing system were:

- costs (low-cost solution);
- motion range (sufficient to implement the assumed research scenarios);
- constant orientation of the end effector of the robot (use of three drives in theory ensured maintaining horizontal position of the working tip, which was a requirement of the designed VR application).

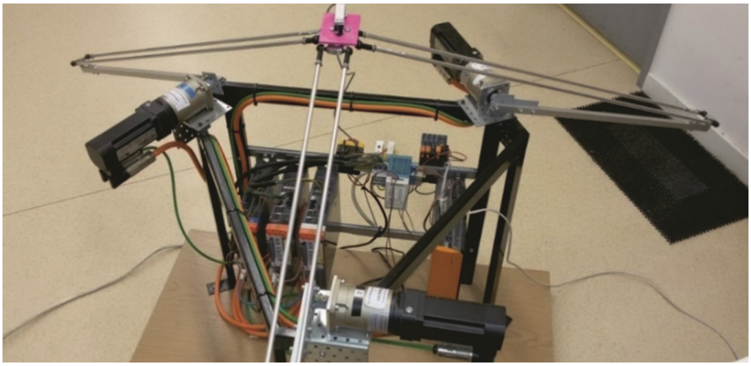


Fig. 1. The Delta robot [27]

The proper configuration [9] of the manipulator, thanks to which it was possible to use it effectively as an active touch device simulating digital objects, was also taken into account.

2.5 Tracked Objects

In addition to the Delta robot, physical models (having their digital representations in VR simulation) were used to deliver the tactile sensation when interacting with elements of the virtual scene. Physical representation of the digital data was crucial from the point of view of the developed VR system for tactile interaction studies. The real, physical models were made on the basis of previously prepared 3D CAD models using additive manufacturing technology of Fused Deposition Modeling (FDM). For the purpose of the experiment, physical models of shafts, a simplified assembly for mounting the shafts and a bracelet enabling tracking of user's hand were made (Fig. 2), they were then implemented in an integrated test stand (equipped with markers to track their position, marker arrangement recorded in the PST system software).

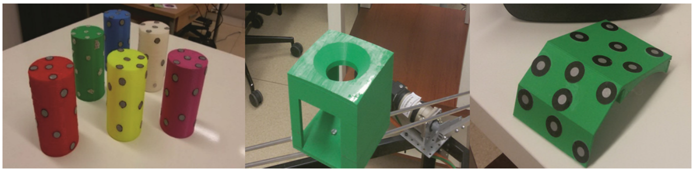


Fig. 2. Physical models used in simulation

2.6 VR Application

The main aim of the research was to check if proposed solutions can be an effective tool for training the activities performed at the production site. The software application was prepared in the Unity 3D programming environment. Bearing in mind the research scenarios (see chapter 3), the following digital models have been placed in the virtual scene (Fig. 3):

- two tables (built from geometric primitives inside the Unity 3D engine);
- shafts and simplified assembly for mounting them (prepared in 3D CAD environment);
- hand (used in the haptic approach)

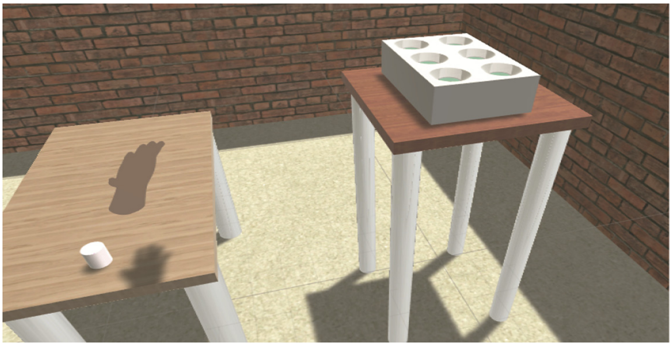


Fig. 3. Objects placed in the virtual scene

Custom scripts have been assigned to both the virtual hand and individual shafts, to allow them to move along their physical counterparts (based on data from the tracking system, sent by the UDP protocol). In each slot of the assembly, a collider object has been assigned. The role of these colliders is to check if shafts were placed in correct slots. The general colliders were placed in the scene, to allow physics-based object interactions with the environment.

3 Test Procedure

3.1 Test Scenario

The experimental research was aimed at checking how much delivery the tactile sensations (with use of various interaction devices) during interaction of a user with elements of the virtual world will affect immersion of the application. In case of the haptic approach, the objects were simulated by the end effector of the touch device and physical models. In case of the non-haptic approach, only the Touch controllers were used. The experiment was a modified “peg-in-a-hole” procedural assembly task (Fig. 4).

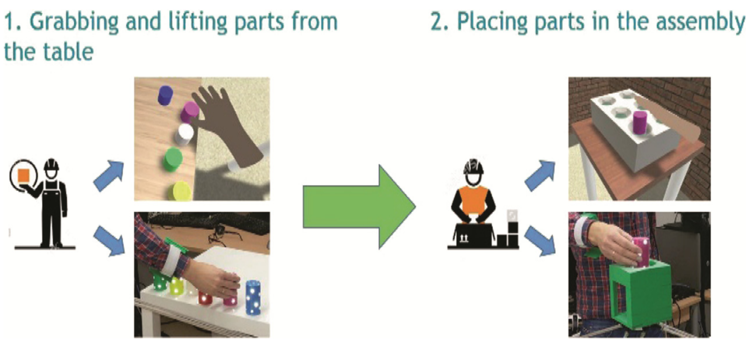


Fig. 4. Procedural task, top – virtual simulation, bottom – physical models

There are 6 shafts in various colors on the virtual table number 1. In the haptic approach, it has its representation in the form of a physical table, on which real six shafts models are placed. On the virtual table number 2, there is a simplified assembly for mounting the shafts (a box with six holes with chamfered edges). In the haptic approach, its physical model with a simplified construction was placed at the end effector of the touch device. It was a real box model with one opening. Task of a user is to move shafts of particular color from the table No. 1 and put them in a specific hole in the assembly. The task of the robot in the haptic approach is to move to a position that allows placing the shaft in a specific hole of the assembly (Fig. 5). In this case, as it is a prototype system, a human operator is necessary- his task is to pull out the physical shaft model from the unit installed at the top of the end effector.

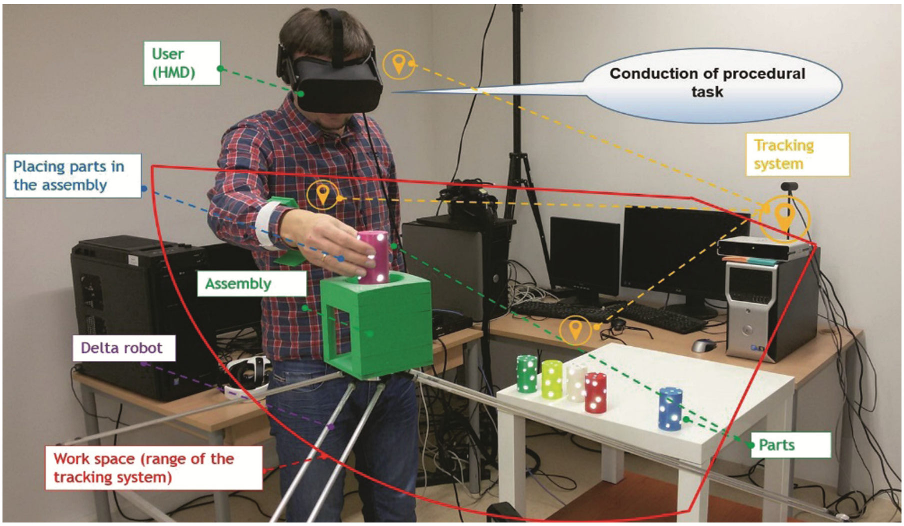


Fig. 5. Research stand for the virtual training of procedural task

In case of the non-haptic approach, based on the Oculus Touch controllers, only the virtual models were used. The user, seeing his virtual hands while holding the controllers is supposed to grab virtual shafts, then press the specific "grasp" button on the controller (a grabbed object is shown in Fig. 6). The next step is to move a virtual shaft to an appropriate position, while avoiding collisions with elements of the virtual environment. After placing a virtual model of the shaft in a correct position, it is necessary to release the button on the controller in order to release the virtual hand and place the shaft in the hole.

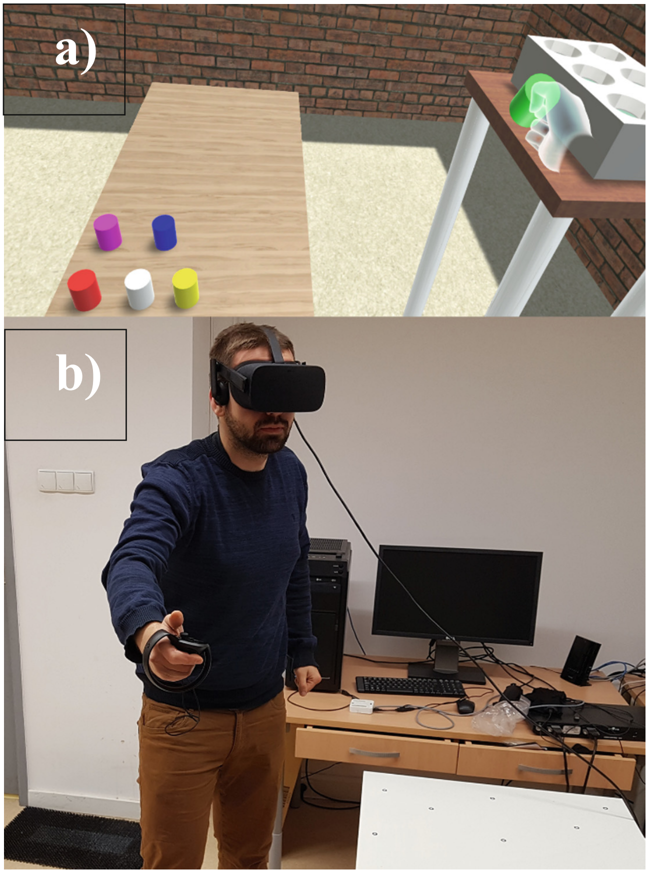


Fig. 6. Grabbed object being transported from Table 1 to Table 2 with Oculus Touch (a - virtual simulation, b – real stand)

Apart from both VR-based approaches (haptic and non-haptic), a physical work stand was also prepared, for the control group (Fig. 7). It consisted of additively manufactured, plastic parts and cardboard elements.

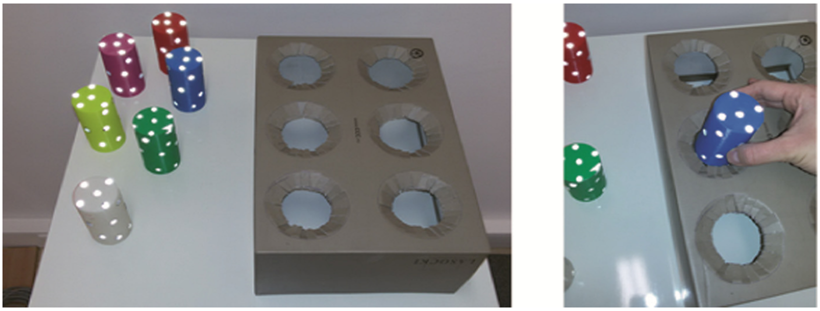


Fig. 7. The physical stand for training

3.2 Test Groups

The tests were aimed at determination of degree of effectiveness of training. Four groups of users were tested (15 persons in each group):

- Group 1 (control group) - persons who have joined the test at the physical stand after 1 min analysis of assembly instructions (the shafts of a particular color in dedicated holes in the assembly were specified) and 4 min of training at the real stand;
- Group 2 - persons commencing the test at the physical stand after 5 min of training completed on the Delta robot and physical models research stand (the haptic approach);
- Group 3 - persons joining the test at the physical stand after 5 min of training using Oculus Touch (the non-haptic approach).

3.3 Collected Data

Each participant in each group carried out five successive attempts to complete the procedural task. The following data was collected for the purpose of evaluating the effectiveness of the training:

- total time of completing the task (in each trial);
- total number of errors (in each sample and total in 5 attempts);
- number of attempts without errors;
- number of attempts with errors;
- number of times when participants asked for help or a graphical hint (in each attempt and in total in 5 attempts).

What was recognized as an error:

- grabbing a part of a particular color in the wrong order;
- placing part of a specific color in the wrong assembly hole.

If a user made an error while performing the experiment, he was immediately informed of that fact by the operator, who recorded the error. At the same time, the participant was informed (in the form of a voice command) which part should be taken (in accordance with the current sequence) or where the part should be placed in the assembly. Hints were displayed in form of graphic prompts (as shown in Fig. 8) - place of the target assembly of the selected part was displayed, depending on the color of the part. In case of such a situation, the time of the task being performed was not stopped.

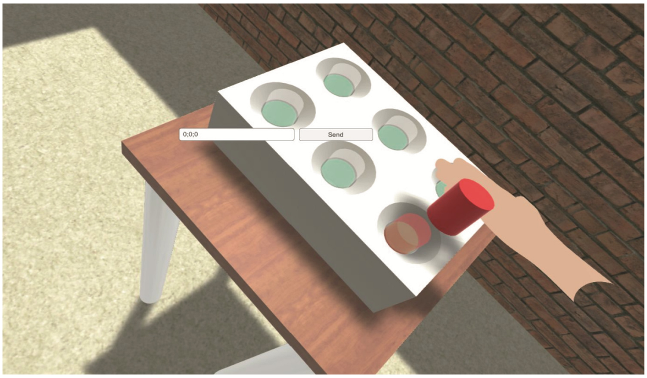


Fig. 8. Graphical helper indicating where the shaft should be placed

After the test at the real stand, the participants of groups 2 and 3 were asked to perform the procedural task one more time in both virtual environments using appropriate sets of devices.

4 Results

Total 45 users took part in the experiments described above. The results (Fig. 9) show that participants of Group 1 performed the best. They completed the procedural task quickly and made the smallest number of errors. The ability to practice on a dedicated physical stand has proved to be a key factor. Worse results were obtained by representatives of Group 2. They completed the task in 10% more time, making more mistakes. Very similar but slightly worse results were obtained by representatives of Group 3. Here, the decisive factor was contact with physical objects - participants knew the correct sequence but for the first time they had to practically put the part in the assembly.

For comparison, representatives of Group 2 and 3 were also asked to perform a measured attempt to perform a procedural task in a virtual environment. Results (Fig. 10) show that using non-contact interaction techniques, users performed the task faster - this may be related to the fact that there was no need to capture physical objects and put them in the assembly, but only “virtual transport” using controllers - restrictions related to precise placement of parts in the assembly were vastly reduced. Users were able to learn the correct sequence, but they did not acquire manual skills. However, when it comes to number of errors and requests for help, the results are very similar. The learning curve of the procedural task changes similarly for all the test groups - progression is observed with each subsequent try.

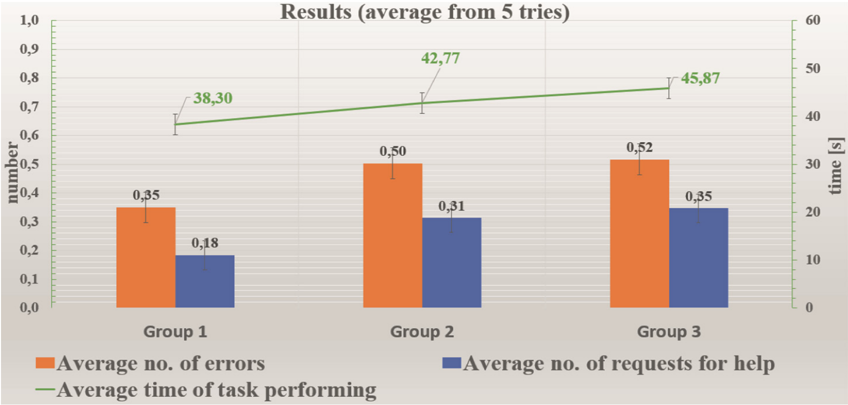


Fig. 9. Results of the test

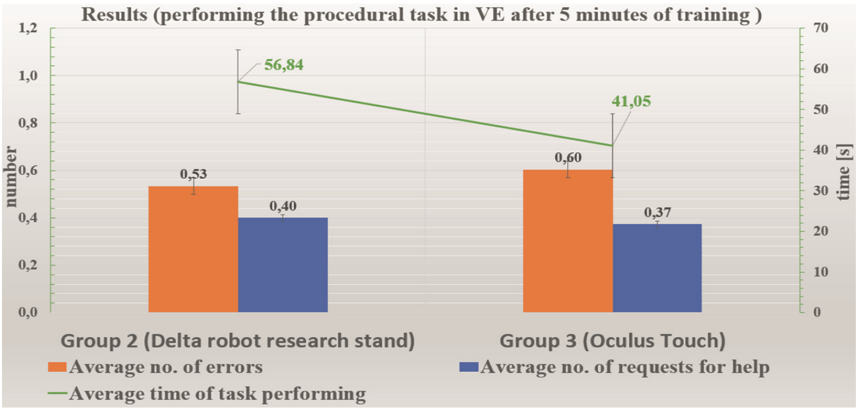


Fig. 10. Performance of procedural task in VE

5 Conclusions

VR solutions are a viable option for training in the field of performing a procedural task. The purpose of the test was to show how the touch affects the immersion of simulation and effectiveness of the learning process of the procedural task. The results are clear - training using both the proposed solutions significantly improves results (fewer mistakes and requests for help) in relation to the group, which had only a traditional instruction. It is worth noting, however, that the discrepancies between the proposed two VR solutions are relatively insignificant. The authors assume that with the increasing complexity of the procedural task, the difference between traditional and virtual training will deepen. It is also worth paying attention to the fact that the Oculus Touch controllers are equipped with motors that enable them to vibrate at selected moments. To a limited extent, this may imitate the tactile impression or at least give additional feedback to a trainee, e.g.

if the device starts to vibrate at the time of collision between virtual elements. In the future, it is planned to examine how increasing the complexity of the task and using the controller's vibrations affect the difference between the proposed methods of VR training.

It is worth noting that the assembled elements were lightweight. In the event that the assembly task would involve parts of significant mass (e.g. if the shafts were made of steel), a person after virtual training (using the Oculus Touch) could feel uncomfortable after starting the operation on a real object. It would result from the difference between the experience gained during the training and the reality. The authors plan to investigate this issue and its impact on the effectiveness of the proposed methods of training as part of further research.

References

1. Martin-Gutierrez, J., et al.: Improving the teaching-learning process of graphic engineering students through strengthening of their spatial skills. Int. J. Eng. Educ. **31**(3), 814–828 (2015)
2. González, M.A., et al.: Virtual worlds. opportunities and challenges in the 21st century. Procedia Comput. Sci. **25**, 330–337 (2013)
3. Berg, L.P., Vance, J.M.: Industry use of virtual reality in product design and manufacturing: a survey. Virtual Reality **21**, 1 (2017). https://doi.org/10.1007/s10055-016-0293-9
4. Hamrol, A., et al.: Virtual 3D atlas of a human body – development of an educational medical software application. Procedia Comput. Sci. **25**, 302–314 (2013)
5. Buń, P., et al.: Educational simulation of medical ultrasound examination. Procedia Comput. Sci. **75**, 186–194 (2015)
6. Grajewski, D., et al.: Improving the skills and knowledge of future designers in the field of ecodesign using virtual reality technologies. Procedia Comput. Sci. **75**, 348–358 (2015)
7. Wu, Y.F., Zhang, Y., Shen, J.W., Peng, T.: The Virtual Reality Applied in Construction Machinery Industry. In: Shumaker, R. (ed.) VAMR 2013. LNCS, vol. 8022, pp. 340–349. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39420-1_36
8. Falah, J., Charissis, V., Khan, S., Chan, W., Alfalah, S.F.M., Harrison, D.K.: Development and evaluation of virtual reality medical training system for anatomy education. In: Arai, K., Kapoor, S., Bhatia, R. (eds.) Intelligent Systems in Science and Information 2014. SCI, vol. 591, pp. 369–383. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14654-6_23
9. Grajewski, D., et al.: Use of delta robot as an active touch device in immersive case scenarios. Procedia Comput. Sci. **104**, 485–492 (2017). https://doi.org/10.1016/j.procs.2017.01.163
10. Bowman, D.A., McMahan, R.P.: Virtual reality: how much immersion is enough? Computer **40**, 7 (2007)
11. Hamblin, C.J.: Transfer of training from virtual reality environments. Ph.D. dissertation, Wichita State University, pp. 1–100 (2005)
12. Gick, M.L., Holyoak, K.J.: The cognitive basis of knowledge transfer. In: Cormier, S.M., Hagman, J.D. (eds.) Transfer of Learning: Contemporary Research and Applications, pp. 9–46. Academic Press, Cambridge (1987)
13. Sowndararajan, A., Wang R., Bowman, D.A.: Quantifying the benefits of immersion for procedural learning. In: Proceedings of the 2008 Workshop on Immersive Projection Technologies/Emerging Display Technologies. ACM, Los Angeles (2008)
14. Lloyd, J., Persaud, N.V., Powell, T.E.: Equivalence of real-world and virtual reality route learning: a pilot study. Cyberpsychology Behav. **12**(4), 423–427 (2009)

15. Breuille, E.: Le simulateur de diagnostic de pannes pour le char Leclerc. Paper presented at the ESAM/NEXTER Forum (2009)
16. Górski, F., et al.: Integracja technik wirtualnej rzeczywistości i wytwarzania przyrostowego – hybrydowe podejście do rozwoju wyrobu. Part 1, Mechanik **3**, 173–176, Part 2, Mechanik **4**, 266–270 (2013)
17. Buń, P., et al.: Low-cost devices used in virtual reality exposure therapy, ICTE 2016. Procedia Comput. Sci. **104**, 445–451 (2017)
18. Grajewski, D., et al.: Immersive and haptic educational simulations of assembly workplace conditions. Procedia Comput. Sci. **75**, 359–368 (2015). https://doi.org/10.1016/j.procs.2015.12.258
19. Gupta, R., Whitney, D.E., Zeltzer, D.: Prototyping and design for assembly analysis using multimodal virtual environments. Comput. Aided Des. **29**(8), 585–597 (1997)
20. Jayaram, S., et al.: Vade: A virtual assembly design environment. IEEE Comput. Graph. Appl. **19**, 44–50 (1999)
21. McDermott, S.D., Bras, B.: Development of a haptically enabled dis/re-assembly simulation environment. In: Proceedings of DETC 1999: ASME Design Engineering Technical Conferences (1999)
22. Ritchie, J.M., Lim, T., Sung, R.S., Corney, J.R., Rea, H.: The analysis of design and manufacturing tasks using haptic and immersive VR - some case studies'. In: Talaba, D., Amditis, A. (eds.) Product Engineering, vol. 5, pp. 507–522. Springer, Dordrecht (2008). https://doi.org/10.1007/978-1-4020-8200-9_27
23. Seth, A., Su, H.J., Vance, J.M.: Development of a dual-handed haptic assembly system: SHARP. J. Comput. Inf. Sci. Eng. **8**, 044502 (2008)
24. www.ps-tech.com
25. Honczarenko, J.: Roboty przemysłowe: budowa i zastosowanie (2009). ISBN 978-83-204-3578-8
26. VISIONAIR project funded by the European Commission under grant agreement 262044, project realised in years 2011–2015
27. www.zum.put.poznan.pl

Study on the Quality of Experience Evaluation Metrics for Astronaut Virtual Training System

Xiangjie Kong(✉), Yuqing Liu, and Ming An

National Key Laboratory of Human Factors Engineering,
China Astronaut Research and Training Center, Beijing, China
kxj2012@outlook.com

Abstract. With the development of virtual reality (VR) technology, it is possible to train astronauts using VR. To make the system more efficient, it is necessary to study the quality of experience (QoE) of astronauts in the virtual environment (VE). Based on the characteristics of virtual training system and the needs of astronauts training, a set of metrics consisting of five higher-level metrics and fifteen lower-level metrics were put forward for the QoE evaluating of the system. In addition, the weight of each higher-level metrics is obtained using analytic hierarchy process (AHP) method. The results of this paper can be used directly in the QoE evaluation of astronaut virtual training system in a quantitative way.

Keywords: Astronaut virtual training system · Quality of experience · Metrics

1 Introduction

Currently, human-centered design is becoming more and more popular. To meet the needs of users is one of the most concerned problems of the developers, and the improvement of the quality of experience has also been paid more and more attention by the developers. The QoE was initially defined as "the overall acceptability of an application or service, as perceived subjectively by an end-user" by the International Telecommunication Union (ITU) in 2007 [1], while the most widely accepted definition was proposed by the Qualinet White Paper in 2012 [2], which is "Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state." This definition has been adopted in 2016 by the International Telecommunication Union [3].

After years of development, virtual reality technology has been applied in various fields, providing a new interactive environment and bringing a totally new experience to the users. We believe VR has very good application prospect with the characteristics of flexibility and safety. In order to provide users with better service and enhance user QoE in virtual environment, it is very necessary to study the quality of experience of virtual reality system.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 416–426, 2018.
https://doi.org/10.1007/978-3-319-91581-4_31

1.1 Related Work

In 2007, the ITU proposed the concept of quality of experience on the basis of quality of service (QoS). Because QoS is a concept in the field of communication, the research on QoE is mainly concentrated in the network and related fields for a long time, such as multimedia streaming services, IPTV, VoIP, interactive network games and so on. As research progresses, the field of QoE becomes more and more broad. At present, the most widely accepted definition of QoE is the definition given by Qualinet in 2012, and the definition has no longer focused on the field of quality of experience, but the user's subjective feelings while using the product or service. As can be seeing from the change of the QoE definition, QoE is very suitable for describing users' subjective feelings about products or services. Therefore, QoE was chosen to evaluate the astronaut virtual training system from the users' point of view.

A great deal of research has been done on the QoE evaluation of the traditional multimedia applications and services, such as video services and audio services. Especially audio services, methods for QoE evaluation of audio services are fully developed. For example, PSQM (Perceptual Speech Quality Measure) and PESQ (Perceptual Evaluation of Speech Quality) proposed by ITU-T P.861 [4] and P.862 [5]. In terms of visual quality assessment, Peak Signal to Noise Ratio (PSNR) is a traditional evaluation index, but it does not take the visual masking phenomenon into consideration. That is to say, every single pixel error affects the value of PSNR, even though the error may not be noticeable. So, a method called MPQM (Moving Pictures Quality Metric) which incorporates human vision characteristics was proposed for video quality assessment [6]. Visual information fidelity (VIF) [7], structure similarity index (SSIM) [8] and texture similarity [9] are several commonly used metrics for visual quality evaluation, which are resulting from the comparison of input and output signals frame-by-frame. Janssen et al. used two metrics, naturalness and colorfulness, to evaluate the visual quality of images and pointed out that people showed a clear preference for more colorful, yet slightly unnatural images [10]. In the FUN model, three metrics, Fidelity, Usefulness, and Naturalness were used to evaluate the visual quality, and the model was considered as a milestone in the road that took visual quality to be evolved into QoE [11]. Zhang and Kuo proposed a model called GLS from the respect of distortion and information loss for the evaluation of QoE on retargeted images and pointed out GLS quality index has stronger correlation with human QoE than other existing objective metrics in retargeted image quality assessment [12]. Three metrics were analyzed in the model. They are global structural distortion (G), local region distortion (L) and loss of salient information (S). In addition, some QoS indicators such as packet loss, delay, jitter, bandwidth, buffer time, and buffer rate are also commonly used for QoE evaluation in network-related applications.

3D images and videos are regarded as milestone in the multimedia technologies, on the other hand, new challenges are introduced in 3D stereoscopic image and video quality assessment compared with traditional multimedia and just a little research has been down on the 3D QoE evaluation. Based on the symmetry of binocular visual perception, Qi et al. proposed a stereoscopic image QoE assessment model [13]. They trained a SVR model for QoE prediction using the HOG features extracted from the two views' image separately. Chen et al. proposed a linear model with three metrics, Image quality, depth quantity and visual comfort, for 3DTV QoE evaluation [14]. Two experiments were designed to determine the weights of the parameters and they are working on a more general model. Similar with Chen's work, Kazuhisa et al. using video quality, depth quality, discomfort and fatigue as the metrics for 3D video QoE evaluation [15]. All their work promotes the study of 3D QoE, while it is not enough.

As the study gets deeper and more extensive, study on QoE is not limited to audio-visual aspects anymore. As described in the Qualinet model, QoE is a multidimensional quality that can be decomposed in a set of perceptual attributes called features, that is perceivable, recognized and namable, and these features can be classified into four categories: features at the level of perception, at the level of interaction, at the level of usage, and at the level of service. As a new media technology, VR technology not only possesses stronger visual expression than traditional media, but also provides new and diversified interaction modes and stronger interaction levels. Some research has been done on VR QoE, for example, Keighrey et al. compared the user QoE of an interactive and immersive speech and language assessment implemented in both AR and VR [16]. Both objective metrics of heart rate and electrodermal activity and subjective metrics of nature of interaction, immersion, discomfort and enjoyment were considered in their study. The findings of the study demonstrate similar QoE ratings for both the AR and VR environments but users acclimatized to the AR environment more quickly than the VR environment. In Hamam's research, five metrics were used for the QoE assessment model, they are Media Synch, Fatigue, Haptic Rendering, Degree of Immersion and User Intuitiveness [17]. Research on VR QoE is just in its infancy and no well-established evaluation methods and standards have been established. Despite the fact QoE is important for both developers and users. Under these circumstances, Huawei initiated the QoE for VR (Quality of Experience for Virtual Reality) project proposal at the ITU-T SG in January 2017, which means a significant step forward for the community in VR QoE evaluation.

1.2 Astronaut Virtual Training System

With the continuous progress of China space station plan, some new needs of astronaut training arose. Problems such as disorientation and the lack of navigation skills in the space station have to be solved urgently. Traditional physical simulators are useful, but they cannot meet all the needs of astronaut training, especially the simulation of micro-gravity environment. VR technology, which has the characteristics of multi-perception, presence, interaction and autonomy, breaks through the limitations of the environment and simulates the micro-gravity environment visually in an immersive way, has the advantages of flexibility and safety, is an alternative and is becoming an important way for astronaut training. The Astronaut Center of China (ACC) has developed an Astronaut Virtual Training System (AVTS) based on VR, which simulates the conceptual structure and internal environment of the Space Station. The AVTS aims to facilitate new and novel experiences for trainees above and beyond what are possible with traditional physical simulators.

Currently, the virtual training system consists of several modules. The scene in the virtual module is shown in Fig. 1.

Fig. 1. Virtual environment in the module

The astronaut virtual reality training system consists of human-computer-interaction interface, simulation software, virtual space environment and astronauts and trainers. The system architecture is shown in Fig. 2. The system uses head-mounted display (HMD), 3D mouse and position tracker as the main interaction equipment. The astronaut virtual training system was designed to conduct training tasks like navigation, instrumentation and objects operation, environment familiarization, extravehicular roaming and so on. During the training process, the trainees sit at the training platform and keep upright. They can control the avatar's movement by the 3D mouse. The data-glove is used to capture the movements of their hands so they can control the avatar's hand movements. With the HMD, trainees can observe the internal environment of the virtual space station by rotating the head.

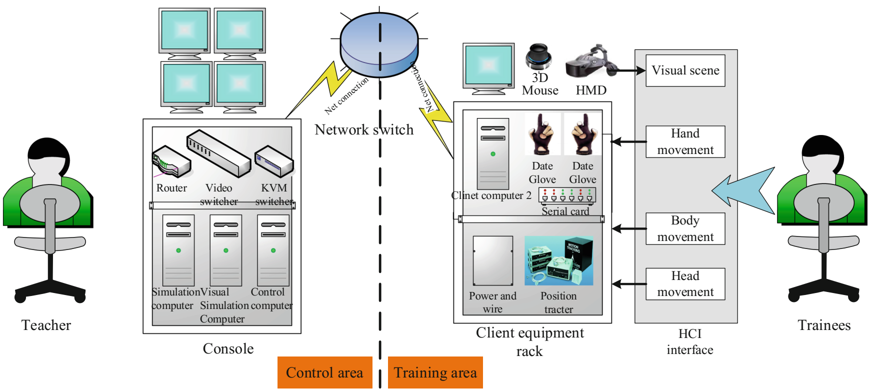


Fig. 2. Astronaut virtual training system architecture

No doubt the VE has its advantages on astronaut training, but the influence of the new technology on astronaut's quality of experience is not considered adequately so far. Based on prior research on QoE and the unique characteristic of the AVTS, a set of metrics is proposed to evaluate the QoE of the AVTS.

2 Method

The easiest and most effective way for QoE measurement is user feedback. However, compared with traditional media quality evaluation, user focuses on more aspects in VE. It is difficult for users to give a reasonable evaluation directly, so the problem needs to be decomposed first. There is no uniform method and metrics for the QoE evaluation until now. Although different organizations established different standards, they all have different focuses. For example, QoE Management Framework (QMF) proposed by ITU-T focused on the evaluation, measurement, basic requirements and preliminary framework of QoE [18], while Customer Experience Management (CEM) framework proposed by TM Forum focused on the definition of CEM, influencing factors and user experience quality model [19]. Obviously, the study on VR QoE is just at its early stage. The existing frameworks or models may not applicable to this new media format and the interactive environment. Coupled with the particularity of astronaut virtual training system, new QoE evaluation metrics need to be built. To the best knowledge of the author, this is the first study on the QoE for a specific VR training system. Based on the definition of QoE and commonly used methods and principles on metrics designing, we proposed five metrics for QoE evaluation, then each metric was further decomposed into multiple lower-level metrics. At the same time, AHP was used to determine the weight of the higher-level metrics.

2.1 Metrics Designing

In order to conduct a reasonable and correct evaluation, we followed four design principles while designing the metrics, which were comprehensiveness, professionalism, rationality and feasibility respectively.

Comprehensiveness means that the study is conducted from different perspectives and the metrics should cover all aspects of the evaluation of QoE. Based on the influence factors of QoE, we evaluate the QoE of AVTS from two aspects of system parameters and user experience. The system parameters refer to the measurable, improved and guaranteed hardware and software characteristics of the system; user experience is the user's subjective experience of the system, including three levels of perception, physiology and psychology.

Professionalism means that the system's specific characteristics must be taken into consideration while metrics designing. In our study, we consider the characteristics not only of VR technology, but also of training system.

Rationality means that the dimension of evaluation is reasonable, and the metrics selection, information collection and coverage range of the metrics have scientific basis.

We put forward the metrics based on existing methods in the literature, and the system characteristics are fully considered.

Feasibility means that the corresponding parameters of the metrics are accessible, and the calculation method is not complicated.

Based on the features of QoE and the application requirements of astronaut virtual training system, a set of metrics for QoE evaluation of astronaut virtual training system are proposed under the consideration of the principle mentioned above. We will evaluate the QoE of AVTS from5 aspects,and they are named functionality, rendering quality, interactivity, side effects and user intuitiveness respectively. Each aspect consists of several lower-level metrics and the lower-level metrics will be obtained through questionnaires. The complete metrics is showed in Table 1, including five higher-level metrics and fifteen lower-level metrics.

Table 1. AVTS QoE evaluation metrics

	The higher-level metrics	The lower-level metrics
QoE	Functionality	Learnability
		Ease of use
		Usefulness
	Rendering quality	Graphics
		Audio
		Cross modality
		Sensory substitution
	Interactivity	Naturalness
		Validity
		Collaboration
	Side effects	Fatigue
		Cybersickness
	User intuitiveness	Presence
		Confusion
		Emotion

2.2 Metrics Interpreting

The interpretation of the metrics is as follows:

Functionality: Functionality is a costumed parameter mainly focus on the acquisition of cognition and skills. It is then subdivided into four parts: Learnability, ease of use, usefulness, and serviceability. Learnability is the measure of the degree to which a user interface can be learned quickly and effectively, that is, the user feels that the interface is familiar and the operation is easy to remember. Ease of use refers to effectiveness at achieving a person's goals in a way that is satisfying to that person, and is fast and error-free. The input and output devices of virtual training system are complex, and task environment of astronaut is quite different from the daily life environment, so there will be more requirements on the users' skills and cognition. It is necessary to design an easy-to-use system so that the astronaut can complete the task better. Usefulness is the extent to which the system actually helps to solve users real, practical problems. A system may be easy to use but not relevant to the actual needs of a user. In our study, the users are expected obtain the appropriate knowledge and skills through training. Here, Learnability and ease of use are two common metrics for system evaluation, while usefulness is a metric put forward with the system characteristics taken into account, that is, the system is designed to train astronauts to acquire relevant skills.

Rendering quality: The rendering quality relates to the quality of the two major modalities, namely: graphics and audio. The visual modality is emphasized because the visual perception is at the dominant position in VE. Moreover, cross-modality and sensory-substitution were also considered. The rendering quality includes not only the system parameters such as the field of view, the resolution and the delay of the HMD, but also the perception parameters such as the feeling of reality and the depth information of the three-dimensional objects in the virtual environment. In addition to visual, auditory information also plays an important role. Although there is a certain tolerance for the non-synchronization of audiovisual, users will be confused if they exceed the threshold. The metric cross-modality is mainly design for the synchronization of different modality. In addition, hearing also plays an important role in sensory-substitution. Without tactile in the VE, visual and auditory information is combined to make reasonable judgement such as whether the avatar has pushed the button, while one can judge this through haptic stimulus in real life.

Interactivity: Interactivity is a major feature of VR, we subdivided this parameter into three parts: naturalness, validity, and collaboration. Naturalness refers to whether the interactive operation is natural and true. The validity refers to whether the result of the operation conforms to the user's expectation, and the cooperation refers to whether the user and the interaction interface can cooperate well. With the use of new interactive devices, the lack of sense (such as tactile) and the existence of sensory-substitution, the interaction in VE is quite different from the traditional way. The way of interaction and the interface must be properly designed so the astronaut can migrate the skills obtained in VE more efficiently to real scene.

Side Effects: This parameter mainly evaluates the fatigue and discomfort caused by the use of the system. Fatigue includes visual fatigue caused by wearing the head mounted display (HMD), muscle fatigue caused by manipulation of 3D mouse and mental fatigue caused by the use of the system; discomfort refers to the symptoms of dizziness, nausea and other simulation sickness. Side effect is a universal problem of virtual reality system. Stanley and Kennedy summarize previous studies and found that 80%–95% of participants had varying degrees of side effects when they were roaming in a virtual environment wearing HMD [20]. In addition, fatigue, especially visual fatigue caused by HMD cannot be avoided.

User Intuitiveness: Intuition reflects the user's most instinctive behavior and feelings, including presence, confusion and emotion. Presence is a unique characteristic of VR system and is a manifestation of the overall performance of the system, which is related to the physical characteristics, virtual presentation technologies, interactive features and the type of task presented by the system. A good sense of presence means that trainees feel themselves in a system-built virtual environment just like in the real world. According to Schmidt and Young's theory of transfer of training [21], if the immersive virtual training environment is similar to the mission environment, virtual training will promote learning and enhance positive transfer, which means that the user's performance in the actual task will be better. Therefore, a good sense of presence not only helps trainees get a better experience, but also helps to improve training performance. Confusion and emotion are the characteristics of the degree of consistency between system response and user expectation. If the operation of the system is inconsistent with the expectation of the user, it will cause confusion; if the system response does not match user's operation, the user will be frustrated and depressed.

2.3 Metrics Weighting

The weight of metrics is the characterization of the degree of importance of metrics. Several methods have been used for metrics weighting in prior research, such as Delphi, AHP, principal component analysis(PCA), artificial neural network and so on.

After the metrics is constructed, the AHP is used to determine the weight of the higher-level metrics. The AHP is a theory of measurement through pairwise comparisons and relies on the judgements of experts to derive priority scales [22]. The comparisons are made using a scale of absolute judgements that represents, how much more, one element dominates another with respect to a given attribute. Table 2 exhibits the scale.

Table 2. The fundamental scale of absolute numbers

Intensity of importance on an absolute scale	Definition	Explanation
1	Equal importance	Two activities contribute equally to the object
3	Moderate importance of one over another	Experience and judgement slightly favour one activity over another
5	Essential or strong importance	Experience and judgement strongly favour one activity over another
7	Very strong or demonstrated importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals	If activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i	
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining n numerical values to span the matrix

Four experts in the fields of virtual reality and human-computer interaction participating in rating the importance of the metrics. Finally, we have the following for the matrix of pairwise comparisons of the metrics with respect to the overall focus, shown in Table 3.

Table 3. Pairwise comparison matrix for higher-level metrics

	Functionality	Rendering quality	Interactivity	Side effects	User intuitiveness
Functionality	1	3	3	2	3
Rendering quality	1/3	1	1	2	2
Interactivity	1/3	1	1	3	2
Side effects	1/2	1/2	1/3	1	1/3
User intuitiveness	1/3	1/2	1/2	3	1

To decide whether the comparison is consistent or not, a consistency test is performed on the judgment matrix. The consistency ratio (CR) helps judge the consistency in pairwise comparison.

For a n-order matrix with the largest eigenvalue of λ_{max}, consistent index (CI) that describes deviation or degree of consistency of the judgment matrix is computed by the following formula:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{1}$$

Then, the consistency ratio (CR) is defined as the following formula:

$$CR = \frac{CI}{RI} \tag{2}$$

Where *RI* is called mean random consistency index, and the value of *RI* is showed in Table 4. The comparison is considered consistently if $CR < 0.1$.

Table 4. Mean random consistency index

Order	1	2	3	4	5	6	7	8	9
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46

The results of the test through calculations showed that $\lambda_{max} = 5.3595, n = 5, CI = 0.0899$ and $CR = 0.080 < 0.1$, so we believe the comparison is consistent.

Finally, the weight of each higher-level metrics through calculations is shown in Table 5.

Table 5. The weight of each higher-level metrics

Functionality	Rendering quality	Interactivity	Side effects	User intuitiveness
0.395	0.180	0.197	0.091	0.136

As shown in the result, functionality has the highest weight of 0.395, which is in line with our expectation. Functionality corresponds to the features of QoE of usage and service. The largest weight also manifests that the function of the astronaut virtual training system is taken seriously. Rendering quality and user intuitiveness corresponding to perception features of QoE, and QoE is characterized from the aspects of system and user respectively. Interactivity is a major aspect of the virtual training system that is different from traditional media. The weight of interactivity is 0.197 and is closed to that of rendering quality, indicating that rendering quality and interactivity are equally important. They two both play important roles in guaranteeing the sense of reality of VE. The weight of side effects and user intuitive were 0.091 and 0.136, respectively, which were relatively low because side effects could be suppressed by reasonably arranging training sessions, and the weight of user intuitiveness should not too large due to individual differences. Therefore, the weight of the index is relatively reasonable. However, due to the fact that the research on the QoE for VR has just begun, there may be some inaccuracies in the expert's experience, and the structure and weight of the metrics will be further optimized in subsequent research, and a set of experiments will be conducted for this work.

In this paper, only the weight of higher-level metrics was calculated. Because the lower-level metrics may be adjusted in practical application, and the processing methods should be different.

3 Conclusion

In this paper, we presented a set of evaluation metrics, consisting of five higher-level metrics and fifteen lower-level metrics, for the QOE evaluation, which makes an evaluation from multiple dimensions. System characteristics and the features of VE are fully considered so that the metrics is applicable for the QoE evaluation of the AVTS. In addition, the weight of each higher-level metrics is obtained through the analytic hierarchy process, which can be used for the QoE evaluation of the system. The result of our study can also be used as a reference for the QoE evaluation of virtual training system in other fields.

References

1. ITU-T: Definition of quality of experience (QoE). International Telecommunication Union, Liaison Statement, Ref: TD 109rev2 (PLEN/12) (2007)
2. Le Callet, P., Möller, S., Perkis, A. (eds.): Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Version 1.2, Lausanne, Switzerland (2013)

3. ITU-T Recommendation P.10: Vocabulary for performance and quality of service, Amendment 5 (2016)
4. ITU-T P.861: Objective quality measurement of telephone-band (300-3400Hz) speech codec (1998)
5. ITU-T P.862: Perceptual evaluation of speech quality (PESQ) (2001)
6. Wang, Y.: Survey of objective video quality measurements. In: International Conference on Intelligent Information Processing, Security and Advanced Communication (2006)
7. Sheikh, H., Bovik, A.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006)
8. Wang, Z., Bovik, A.C., Sheikh, H.R., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
9. Zujovic, J., Pappas, T.N., Neuhoff, D.L.: Structural texture similarity metrics for image analysis and retrieval. IEEE Trans. Image Process. Publ. IEEE Sig. Process. Soc. **22**(7), 2545–2558 (2013)
10. Janssen, T.J.W.M., Blommaert, F.J.J.: Image quality semantics. J. Imaging Sci. **41**(5), 555–560(6) (1997)
11. Ridder, H.D., Endrikhovski, S.: image quality is fun: reflections on fidelity, usefulness and naturalness. In: SID Symposium Digest of Technical Papers, vol. 33, no. 1, pp. 986–989 (2002)
12. Zhang, J., Kuo, C.C.J.: An objective quality of experience (QoE) assessment index for retargeted images. In: Proceedings of ACM International Conference on Multimedia. ACM (2014)
13. Qi, F., Zhao, D.: Quality of experience assessment for stereoscopic image based on the symmetry of binocular visual perception. Intell. Comput. Appl. **6**(4), 72–74 (2016)
14. Chen, W., Fournier, J., Barkowsky, M., Callet, P.L.: Quality of experience model for 3DTV. IS&T/SPIE Electron. Imaging **8288**, 2978–2982 (2012). International Society for Optics and Photonics
15. Kazuhisa, Y., Taichi, K., Kimiko, K.: QoE assessment methodologies for 3D video services. NTT Tech. Rev. **11**(5), 1–5 (2013)
16. Keighrey, C., Flynn, R., Murray, S., Murray, N.: A QoE evaluation of immersive augmented and virtual reality speech & language assessment applications. In: Ninth International Conference on Quality of Multimedia Experience. IEEE (2017)
17. Hamam, A., Saddik, A.E.: Evaluating the quality of experience of haptic-based applications through mathematical modeling. In: IEEE International Workshop on Haptic Audio Visual Environments and Games, pp. 56–61. IEEE (2012)
18. ITU-T: ITU-T Recommendation P.10/G.100 "Amendment 2: New Definitions for Inclusion in Recommendation ITU-T P.10/G.100". International Telecommunication Union, Geneva, Switzerland (2008)
19. Customer Experience Management: Driving Loyalty and Profitability TMForum, NJ, USA (2008)
20. Stanney, K.M., Kennedy, R.: Simulation sickness. In: Vincenzi, D.A., Wise, J.A., Moulous, M., Hancock, P.A. (eds.) Human Factors in Simulation and Training, pp. 117–128. CRC Press, Boca Raton (2008)
21. Schmidt, R.A., Young, D.E:. Transfer of movement control in motor skill learning. Transf. Learn. Contemp. Res. Appl. 47–79 (1987)
22. Saaty, T.L.: Decision making with the analytic hierarchy process. Int. J. Serv. Sci. **1**(1), 83–98 (2008)

Developing and Training Multi-gestural Prosthetic Arms

Albert Manero[1], John Sparkman[1], Matt Dombrowski[2], Ryan Buyssens[2], and Peter A. Smith[2(✉)]

[1] Limbitless Solutions, Orlando, USA
{albert,john.sparkman}@limbitless-solutions.org
[2] University of Central Florida SVAD, Orlando, USA
{MattD,Ryan.Buyssens,Peter.smith}@ucf.edu

Abstract. Learning to use prosthetic limbs is challenging for children. One solution to this is to design engaging training games that teach children how to use their new limbs without boring or fatiguing them. The interdisciplinary design team, comprised of digital media faculty, researchers, engineers, and health professionals, strives to create innovative solutions. The training program will provide recipients with the opportunity to become proficient at using their prosthetics to ensure their successful, long term use of these limbs. This collaboration among physical trainers, engineers, and psychologists will create fun, kid-friendly training solutions that follow sound physical training guidance.

Keywords: Games · Virtual reality · Prosthetics · Alternate control

1 Introduction

The recent popularity of 3D printed technology to create prosthetic limbs is not surprising [1]. These technologies drastically reduce the cost of prosthetic arms making them available to children, a population that often time had to go without a prosthetic, until they stopped growing. This left them with a lifetime of not using a prosthetic. As a result, several groups have formed to help bring arms to children who need them [2]. One such group is Limbitless Solutions, a non-profit organization at the University of Central Florida (UCF) dedicated creating personalized prostheses for children with disabilities [3].

The mission of Limbitless Solutions if different. The organization not only creates innovative arms, but also incorporates the interests of the children receiving them into their design. Children can get arms that look like super heroes, magical creatures, video game characters, or anything the like; making the arm not just a prosthetic but an extension of their personality.

As a new generation of arms is being designed at Limbitless Solutions there is also a need to train the children to use them. In an effort to find a fun way to incorporate training and real-time feedback a game based training system has been developed.

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 427–437, 2018.
https://doi.org/10.1007/978-3-319-91581-4_32

2 Background

As of 2005 the number of cases of limb loss in the United States had increased to 1.6 million [4], with projections expecting these numbers to advance to over 3.5 million by 2050. A variety of factors are driving this rapid increase including dysvascular conditions (which in part can be due to diabetes), trauma events, cancer cases, and congenital amputations (Health Care Utilization Project National Inpatient Sample (HCUP-NIS), 1996. Studies have shown that underrepresented minorities have risk factors of 2 to 3 times as likely as non-Hispanic [5].

Congenital upper limb differences occur approximately 2.6 of every 10,000 live births [6] and impacts upper limbs at 1.6 times as frequently as lower limbs. For children, congenital birth defects are the leading cause of limb differences [7]. There remain several challenges for prosthetics adaptation (costs, children's growth rates, insurance and health care policies just to name a few), and the limited extent of adoption has an impact on both the populations' functionality (both children and adults) as well as their community participation. However, over 85% of artificial limbs worn are for lower limbs as the accessibility and adoption for upper limbs has far lagged behind [8].

Quality of life scoring for people with limb difference has helped to identify different aspects of the challenges associated with upper limb differences including: physical disability, pain, energy, emotional reaction, social isolation, and quality of sleep [9]. For both male and female persons, emotional reaction and social isolation scored the lowest quality of life findings, even beyond the physical limitations. This has also been observed with the pilot study of bionic kids that have worked on the prototype development here at Limbitless Solutions at the University of Central Florida. These psychosocial factors interact with children's development progression (Psychological adjustment and perceived social support in children with congenital/acquired limb deficiencies. Journal of behavioral medicine) and can be a burden. Social support plays a large role in Lazarus' cognitive model [10], and children with limb deficiencies have linked deficiencies with depressive symptomatology, trait anxiety, and general self-esteem. In the words of our bionic kids, Alex was afraid to go to the grocery store because of how people would state or speak to him due to his limb difference.

Functionality, children are highly adaptable and are observed to overcome nearly any challenge [7] on a day to day basis. However, these challenges cause substantial efforts and potential fatigue points in everyday tasks [11]. From the experience with our bionic kids, using their bionic arm to support their bike riding they were able to bike further with less physical strain for longer durations of time, whereby being able to engage in more meaningful engagement with the peers that may lead to a stronger support system and psychosocial development. Zachary's (one of our bionic kids) family described to our team how the need to contort (Zachary's) spine to have his limb difference arm reach the handle-bars limited the time he could spend in the activity and contributed in him being left behind during activities with his friends.

Further, younger children demonstrate higher degrees of usefulness with a prosthesis as compared to older children [11]. Our mission is to provide children and younger adults a functional and expressive electromyographic actuated arm experience, by increasing access to the technology that can also grow with them as needed via 3D printing modular components.

A major challenge that may limit the adoption of electromyographic bionics, particularly for children, is learning and adapting to the control schemes (Surface EMG in advanced hand prosthetics). Active controls for multiple gestures have remained complex and include machine learning techniques [12, 13] but have larger computational requirements. This may be a limiting factor towards the adoptability for children to have multi-gesture controls. More recently, an effort has been made to gamify the learning of prosthetics control [14]. This has provided feedback that better links the stimulus (a muscle flex) with a potential set of outcome engagements on a prosthetic.

Our team has taken this a step further and has created multi-gesture electromyographically triggered video games that utilize a special electromyographic controller. This has produced a study to capture the engagement of the training games for children, with excellent affinities reported thus far. In our program, a single electromyographic sensor has been employed with multi-gesture capability with short pattern recognition and magnitude control, that has been simplistic enough for children to quickly grasp and engage. This type of engagement is believed to encourage stronger affinities and subsequent adoption of the device for children and young adults, which is believed to correlate to stronger social integration, spatial (dexterity) manipulation, and community engagement. These factors correlate to long term success from the classroom, to the career field, or in the home.

3 Current State of Development

The current state of the technology has produce a fully functional prototype for both the bionic limbs and video game training system featured here in Fig. 1. The bionic arm features custom circuitry, professionally manufactured, electromyographic binary actuation for hand open and close. Battery life currently averages approximately 8 h and can be charged with equipment similar to a laptop charger. The video game training system includes a custom electromyographic actuated wearable controller, that connects via USB to a computer, and several games have been produced and demoed for children with and without limb deficiency. The system has been well received and survey information is currently being collected for evaluation.

Fig. 1. Bionic arm prototype (left) and video game training system (right)

Fundamental technology improvements to the product are for development. These include improving both the (i) dexterity and (ii) degrees of actuation. The assessments proposed are designed to give feedback on the necessity of the refinements as well as the refinements implementation, across the course of the five-year study. Multi-finger actuation dexterity is of primary importance. This includes functional gestures, for example pinch grip. The current state of the prototype hardware allows for individual finger controls, but refinement to the software for assessing intentionality and cycling through gestures is still in development. Examples of the upgrade hardware are presented in Fig. 2. The training software protocols are designed to teach coordinated flexing patterns and magnitudes, while also assessing the ease of capturing intentionality of the user.

Fig. 2. Multi-gesture bionic hardware design on right with exploded view on left

Also in development is an electromyographic mechatronic elbow. This, as is the case with the multi-finger actuation, can be achieved with only a single electromyographic sensor via gesture analysis. This proof of concept has been completed and tested in our prior work, and allows for a reduction in weight and hardware complexity which benefit the user's experience greatly. The mechatronic elbow will substantially improve the functionality of the above-elbow congenital amputee population, but requires the novel training programs to ensure successful adoption particularly for children.

3.1 Video Game Prototypes

Learning to use prosthetic limbs is challenging for children. A portion of the proposed funding is to design engaging training games that teach children how to use their new limbs without boring or fatiguing them. The interdisciplinary design team, comprised of digital media faculty, researchers, engineers, and health professionals, strives to create innovative solutions. The training program will provide recipients with the opportunity to become proficient at using their prosthetics to ensure their successful, long term use of these limbs. This collaboration among physical trainers, engineers, and psychologists will create fun, kid-friendly training solutions that follow sound physical training guidance. The Limbitless Solutions game design team has already developed game controllers that utilize electromyographic sensors bridging the 3D printed prosthetic multi-gesture functions to the training games. The award-winning beta versions of these games were created from seed funding and support from the University of Central Florida (UCF). The prototype games were featured at the Smithsonian American Art Museum in Washington D.C. and the Game Developers Conference (GDC) in San Francisco. These prototypes succesfully showcased hybrid multidisciplinary interactive experiences that deliver both arts, training, and cultural impact.

Our interdisciplinary team is comprised of game design and digital media professors, researchers, and visual artists, engineers, and health professionals UCF. We seek to further evolve these individual games into a multi-faceted training program for limb recipients. The developed games will train the user by establishing a multi-phase training system varying in difficulty. Games within each phase will train different gesture types: (i) grasping and releasing; (ii) pre-set finger combinations (thumbs up, peace sign, etc.); and (iii) individual finger dexterity.

The multi-phase system, drawing from the strengths of such a variety of disciplines and fields of knowledge, will integrate encouragement, humor, and, most importantly, fun, to provide a safe and healthy training experience for children learning to use their 3D printed prosthetic limbs, placing them on the pathway to a life of increased capabilities and success.

3.2 STEAM Education: Internal and External

The merging of arts, engineering, and health care is pivotal to the success of our project. Our game design teams are composed of over 25 current Science, Technology, Engineering, Arts, and Math (STEAM) undergraduate, graduate, and doctoral students. The melding of disciplines and mindsets allows true cross-discipline innovation to occur. Externally it is important for the team to also work with their client base. The game design team receives direct input on art content and narrative from the recipients of the bionic arms. The team also works closely with area elementary schools in showcasing the impact art and engineering games have on the community, and has recently become a destination for local schools to visit as part of an engagement field trip for STEAM learning.

4 Evaluation Methods of a Prosthetic Arm

Prosthetic arms for children provide an impressive number of challenges. It is important to understand exactly how they are working and how to make them work better for the children. The same can also be said for the games research. There are a variety of assessment vehicles for quantifying the impact and use of prosthetics and video games; several methods are presented herein.

4.1 Assessment of Capacity for Myoelectric Control (ACMC)

This assessment for myoelectric device control is a Rasch rating scale that is used to detect expected change in a person's ability using objective variables. There are 30 items that evaluate a prosthetic arm's ability to do specific functions that involve gripping, holding, releasing, and coordinating between limbs. This is done by asking the subject to perform certain tasks and scoring their motions by a trained professional [15].

4.2 Unilateral Below Elbow Test (UBET)

UBET was designed to evaluate the functionality of individuals both wearing and without wearing a prosthesis and is comprised of nine age appropriate tasks. Two scales for quantified scoring including Completion of Task and Method of Use are used to evaluate a child's function and adaptability on a 5 point scale. These scales range from 0 [unable to complete the task] to 4 [completes task without difficulty]. This analysis method allows for a direct comparison of the individual with and without their prosthetic device for measuring direct function improvements [16].

4.3 Pediatric Outcomes Data Collection Instrument (PODC)

PODC is a well-validated musculoskeletal health questionnaire, and published results of this measure are available for the general population and for a normal population. PODCI questions address both the activity and the participation components of function. The PODCI contains six validated scales, or domains (Upper Extremity Physical Function; Mobility/Transfers; Sports/Physical Function; Pain/Comfort; Happiness; and Global Function, which is a combination of Upper Extremity Physical Function, Mobility/Transfers, and Sports/Physical Function), with a total of 108 questions. The maximum score in any domain is 100 points, and scores below the low 80 s are considered to represent below-normal function. Parents answer for children who are two to ten years of age, and children and parents answer for those who are eleven to twenty years of age [17].

4.4 Pediatric Quality of Life (PedsQL)

PedsQL is a well-validated survey that asks twenty-three questions of both parents and children about various aspects of health-related quality of life over the past month. Published results are available for the general population. The PedsQL scoring algorithm translates the available responses to questions ("never," "almost never," "sometimes," "often," or "almost always") into scores of 0%, 25%, 50%, 75%, and a maximum of 100% for each of four generic core scales (Physical Health, Emotional Functioning, Social Functioning, and School Functioning). The raw data used to obtain Emotional, Social, and School Functioning scores are averaged to obtain the Psychosocial Health Domain score, and the raw data used to obtain all four core-scale scores are averaged to obtain a Total Scale Score. Parents answer for children who are two to four years of age, and both the parents and the children answer for those who are five to twenty years of age [18].

4.5 Prosthetic Upper Extremity Functional Index (PUFI)

PUFI is effective at measuring the extent prosthetic limb use, the challenge or ease of use, and overall task performance as reported by a parent or the individual. It categorizes real life tasks by age group for evaluation that can include for example: zipping a zipper, grasping both handles of a bicycle, or peeling a banana just to name a few. In addition to task performance completion, the ease of use is measured for a more comprehensive assessment [11].

4.6 System Usability Scale (SUS)

SUS is an effective survey for measuring both the usability and context for video games. It is based on a Likert scale, with degrees of agreement and disagreement on a 5 point scale. The composite score is ranged from 0 to 100, and has been used in the past to measure and contrast both the software as well as the hardware systems [19].

4.7 Game User Experience and Satisfaction Survey (GUESS)

GUESS this survey helps to measure the engagement of a video game for a group of users, and is usually scored at the end of their play-testing. It includes categories such as: engagement, immersion, flow, presence, and existing gaming scales. This survey vehicle is critical for evaluating a variety of games for a variety of participants, and helping to determine the resonance and play-ability of these in-house developed training games [20].

5 A Game Design to Support Multiple Gestures

Disabled video gamers are an often overlooked community in the gaming world. In 2008, Information Solutions Group surveyed over 13,000 casual gamers. In that survey 20.5% identified as disabled gamers. Furthermore, more than 10% have stated that they have had casual games recommended by a doctor; The survey also deemed that disabled gamers play longer and more than non-disabled gamers. [21] The team previous experience of accessibility forward game design that has been showcased on an international academic level. The games research team is composed of researchers in the fields of game design, visual arts, and applied design. The games team has already successfully created and implemented EMG training controllers that interface with the 3D printed prosthetic arm of Limbitless Solutions.

These initial games utilize an interface that allows for a thresholded input. That is when a user is a rest the EMG will generate a lower value then when the muscle is flexed, and while this number can be different between users a simple calibration allows the user's flex to be read as a single button input to a video game. Early games would use the thresholded input to shoot a gun on a space ship, manipulate a finger picking a nose, have a dog jump a fence, and many other fun activities that provided instant feedback to the player. This input worked very similarly to the input of the prosthetic arm, which opens and closes like a garage door opener on flex. Figure 3 shows a rendering of the EMG controller on the left and a screen shot of Beeline Border Collie on the right.

Beeline Border Collie is very similar to the popular game Flappy Bird. These simple games provide a great introduction to how the prosthetic arms work, and allow for exercising prior to receiving a prosthetic arm. But as the arms improve in quality the need to train more complex behaviors is needed. These games need to consider the input

Fig. 3. Rendering of the custom EMG game controller left and Beeline Border Collie right

needed to manipulate a more complex arm with gestural controls and the ability to individually articulate fingers. In an effort to explore this control the game Magical Savior of Friends was born.

Magical Savior of Friends follows a small elf like magician that is helping a group of frogs fight off an evil group of snakes. The snakes are led by Sir Snikelsworth, a greedy landlord that has raised the rent on the frogs to the point that they can no longer afford to live in the kingdom. The game provides a rich platforming experience, similar to what might be found in a Mario Brothers game, but with one major difference. The player can control magic based on how they flex.

Using a proprietary celebration technique, the game can not only read if the player is flexing, but can also determine intensely they are flexing, and for how long. In the game this allows for 3 types of lightning magic when using arcane magic. This might be called multi-thresholded input. It allows players to throw a rock at various angles using earth magic. This is a truly analog input. Water magic requires multiple flexes to control. Fire requires sustained flexing. Air power requires sustaining and thresholding to fly around like in Flappy Bird. These magics can be combined in complex ways and each will represent a real function in the prosthetic arm. This game is training players to flex patterns that will empower gestures or individual finger articulations in the future.

Figure 4 shows the Magical Savior standing on a platform near one of the sage frogs. A collectable coin is seen on the edge of the screen. There is also an imprisoned frog on the far right. These frogs can be found hidden in levels and freed. In the upper left is the flex meter. It looks like a hand and shows the player how much they are flexing at any given point in the game.

Fig. 4. A screen shot of magical savior of friends

6 Conclusions

The ability to use games for training with this child based population is incredibly powerful. The children are engaged in the training and have created fan art and stories around the characters in the games. The development team has incorporated this feedback into the games and so the games feel custom built, but also personally designed by the children. Making learning a fun and rewarding, has allowed for more complex behaviors to be mixed in quickly. Also, the games platform can act as a prototyping platform for new ideas in the future.

6.1 Future Work

As the user progresses through the training game, they will unlock the use of different in game abilities. These abilities will directly coordinate with the controls and rhythmic patterns needed to control multiple dexterous gestures in the prosthetic arm. The games and controllers utilize a custom built electromyographic game controller that can be calibrated to the individual user. Typically, without the games, it is difficult to communicate to a child the percentage of flex required to operate a specific gesture for the bionic arm. Furthermore, it becomes a challenge to then incorporate multiple flex levels and patterns and to be able to differentiate them from other commands. The games calibration system also acts as a visual cue to distinguish different magnitude levels of flex. This allows the research team to map those different levels of flex to specific gestures for the arm. In the game play each level of flex will be mapped to a distinct mechanic that is pivotal for the games success. For our bionic kids who have tried the beta training games, we have seen immediate results in gesture learning and engagement.

References

1. Michalski, M.H., Ross, J.S.: The shape of things to come: 3D printing in medicine. JAMA **312**(21), 2213–2214 (2014)
2. Tanaka, K.S., Lightdale-Miric, N.: Advances in 3D-printed pediatric prostheses for upper extremity differences. J. Bone Joint Surg. Am. **98**(15), 1320–1326 (2016)
3. Limbitless Solutions (2018). http://limbitless-solutions.org/about-us/. Accessed 06 Apr 2018
4. Ziegler-Graham, K., MacKenzie, E.J., Ephraim, P.L., Travison, T.G., Brookmeyer, R.: Estimating the prevalence of limb loss in the United States: 2005 to 2050. Arch. Phys. Med. Rehabil. **89**(3), 422–429 (2008)
5. Lavery, L.A., Ashry, H.R., Van Houtum, W., Pugh, J.A., Harkless, L.B., Basu, S.: Variation in the incidence and proportion of diabetes-related amputations in minorities. Diabetes Care **19**(1), 48–52 (1996)
6. Shen, Y., Drum, M., Roth, S.: The prevalence of perioperative visual loss in the United States: a 10-year study from 1996 to 2005 of spinal, orthopedic, cardiac, and general surgery. Anesth. Analg. **109**(5), 1534–1545 (2009)
7. Yiğiter, K., Şener, G., Erbahceci, F., Bayar, K., Ülger, Ö.G., Akdoğan, S.: A comparison of traditional prosthetic training versus proprioceptive neuromuscular facilitation resistive gait training with trans-femoral amputees. Prosthet. Orthot. Int. **26**(3), 213–217 (2002)
8. NLLIC and the Limb Loss Research and Statistics Program: Limb Loss in the United States, Amputee Coalition of America (2007). http://www.amputee-coalition.org/fact_sheets/limbloss_us.html. Accessed 06 Apr 2018
9. Stucki, G., Ewert, T., Cieza, A.: Value and application of the ICF in rehabilitation medicine. Disabil. Rehabil. **24**(17), 932–938 (2002)
10. Lazarus, R.S., Folkman, S.: Coping and adaptation. In: The Handbook of Behavioral Medicine, pp. 282–325 (1984)
11. Wright, F.V., Hubbard, S., Naumann, S., Jutai, J.: Evaluation of the validity of the prosthetic upper extremity functional index for children. Arch. Phys. Med. Rehabil. **84**(4), 518–527 (2003)
12. Ekvall, S., Kragic, D.: Grasp recognition for programming by demonstration. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, pp. 748–753(2005)
13. Bitzer, S., Van Der Smagt, P.: Learning EMG control of a robotic hand: towards active prostheses. In: Proceedings of the 2006 IEEE International Conference on Robotics and Automation, ICRA 2006, pp. 2819–2823 (2006)
14. Anderson, F., Bischof, W.F.: Augmented reality improves myoelectric prosthesis training. Int. J. Disabil. Hum. Dev. **13**(3), 349–354 (2014)
15. Lindner, H.Y., Langius-Eklof, A., Hermansson, L.M.: Test-retest reliability and rater agreements of assessment of capacity for myoelectric control version 2.0. J. Rehabil. Res. Dev. **51**(4), 635–644 (2014)
16. Bagley, A.M., Molitor, F., Wagner, L.V., Tomhave, W., James, M.A.: The unilateral below elbow test: a function test for children with unilateral congenital below elbow deficiency. Dev. Med. Child Neurol. **48**(7), 569–575 (2006)
17. Raichle, K.A., Hanley, M.A., Molton, I., Kadel, N.J., Campbell, K., Phelps, E., Ehde, D., Smith, D.G.: Prosthesis use in persons with lower-and upper-limb amputation. J. Rehabil. Res. Dev. **45**(7), 961 (2008)
18. Varni, J.W., Seid, M., Kurtin, P.S.: PedsQL™ 4.0: reliability and validity of the pediatric quality of life inventory™ version 4.0 generic core scales in healthy and patient populations. Med. Care **39**(8), 800–812 (2001)

19. Brooke, J.: SUS-a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
20. Phan, M.H., Keebler, J.R., Chaparro, B.S.: The development and validation of the game user experience satisfaction scale (GUESS). Hum. Factors **58**(8), 1217–1247 (2016)
21. Information Solutions Group: Survey: 'Disabled Gamers' Comprise 20% of Casual Games Audience. Cision PR Newswire (2008). https://www.prnewswire.com/news-releases/survey-disabled-gamers-comprise-20-of-casual-video-games-audience-57442172.html

Virtual Reality Based Space Operations – A Study of ESA's Potential for VR Based Training and Simulation

Manuel Olbrich[1(✉)], Holger Graf[1], Jens Keil[2], Rüdiger Gad[3], Steffen Bamfaste[4], and Frank Nicolini[5]

[1] Fraunhofer IGD, Fraunhoferstrasse 5, Darmstadt, Germany
{Manuel.Olbrich,Holger.Graf}@igd.fraunhofer.de
[2] Visometry GmbH, Fraunhoferstrasse 5, Darmstadt, Germany
Jens.Keil@visometry.com
[3] Terma GmbH, Europaplatz 4, Darmstadt, Germany
ruga@terma.com
[4] ESA/ESOC, European Space Operation Centre, Darmstadt, Germany
Steffen.Bamfaste@esa.int
[5] EAC - European Astronaut Centre, Cologne, Germany
Frank.Nicolini@esa.int

Abstract. This paper presents the results of a study the authors conducted together over a year in order to identify key issues of ESA's (European Space Agency) potential for a deployment of Virtual Reality training environments within space operations. Typically, ESA simulates several operations using DES like systems that need to be linked to a VR environment for training purposes. Based on the second generation of VR equipment and development tools the paper describes a holistic design approach from scenario development through design decisions on SW and HW choices until the final development of a PoC for a virtual lunar base that might simulate the metabolism of a lunar base. Here the idea was to mirror the mass- and energy-flows within a lunar base in order to maintain an environment, in which astronauts can live and work and to establish a tool that supports the training of astronauts for operating such a lunar base, the one likely next step of human space exploration beyond the International Space Station as identified by ESAs decision makers. In the end, we have realized a PoC for a fire emergency case on a lunar base allowing astronauts being trained in a fully simulated and integrated environment. The system could be tested and evaluated in two set-ups, first using classical VR controllers, second, using recent VR glove technology.

Keywords: Virtual reality · Glove control · Virtual lunar base · DES coupling

1 Motivation

"There's life in the old dog yet" might describe the "second life" of Virtual Reality (VR) based systems and thus several new endeavors for the realization of VR based training and simulation systems. In the last 3–4 years, we have been facing a second wave of VR technology conquering consumer markets and industrial applications. This is, by no

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 438–451, 2018.
https://doi.org/10.1007/978-3-319-91581-4_33

means, only been pushed by new hardware and affordable pricing models, but also new technology stacks, that have been designed and developed specifically for interactive applications being deployable on any potential device or platform. Coming with the miniaturization of sensory and its modalities, new ways of perceiving the environment, new forms of interaction and new ways to distribute and deploy interactive content have been realized. Here, the precision and easiness of technology deployment might have been a third pushing factor for the revival of VR based systems, leading to a new hype of its use. The potential of precise, low cost VR and new forms of HCI also inspired new approaches for training and simulation environments. Therefore, the European Space Agency (ESA) started an activity to establish a study on the deployment of VR technology for Space Operations, its training and coupling to the internal simulation backend. This paper presents a holistic approach for the design and realization of a Virtual Reality (VR) based workplace that enables ESAs decision makers, astronauts or operators to conduct interactive explorations of different situations using recent technologies in view of the realization of a virtual lunar base.

2 Background

ESA identified “lunar exploration” as one likely next step of human space exploration beyond the International Space Station (ISS), which operates in Low Earth Orbit. One option is the establishment of a crewed lunar base for long-duration human presence on the Moon. For studying and optimizing the metabolism of a lunar base (i.e. the mass- and energy-flows within the lunar base in order to maintain an environment, in which astronauts can live and work) and to establish a tool that supports the training of astronauts for operating such a lunar base, the use of a virtual lunar base appeared to be a promising option [1]. In this scenario, one major challenge is the coupling of ESA’s simulation backend to a real-time responsive, immersive VR environment. Typically, the simulation backend offers services based on discrete event simulation models (DES) that provide access to run-time data of space systems (Space Systems Simulators Infrastructure (SIMULUS)), or functions and services specifically serving the needs of spacecraft monitoring and control systems (MCS - Mission Operations Support Infrastructure (MICONYS)). Although many of the DES systems do offer the possibility of interaction, they lack the ability to place the user of the simulator into the center of operations that might be realized through an immersive 3D representation of a simulated scenario [2]. The potential to directly link operations simulation to an immersive virtual reality (VR) environment, allowing users to interactively change the simulation while in process, opens new ways for exploring complex interactions between model users, objects and operations being simulated. In previous works, [3] examined the promise, at that point in time, VR presented to developers of DES models, though initial enthusiasm has been thwarted by the tremendous overhead in 3D content creation and the limitations of VR. Nevertheless, the authors established the notion of VR-based DES or VRSIM. A comprehensive overview on VRSIM or VR-DES technologies is given in [4], concluding that mainly industrial manufacturing systems in the view of smart factories have been realized as testbeds for VR environments with a focus on the whole value chain optimization for rapid decision making in simulated “what-if scenarios”. The authors also identified future endeavors that will be focusing on new sensory equipped and networked environment with many more real-time (big) data that might need to be pushed into a VR environment. Although many best-practices have been established [4], any coupling to existing infrastructure components comes with its own peculiarities and specific problems imposing new challenges on VR based resp. “user-in-the-loop” exploration of DES generated data.

3 Overall Approach

3.1 Ideation and Scenario Development

We were starting our study with ideation and concept creation sessions with several ESA stakeholders involved (i.e. ESA’s space operation center – ESOC, ESA‘s astronaut training center – EAC, and ESA’s research and technology center – ESTEC), to draw on different scenarios of use and derived use cases. The findings should provide the basis on the choice of technology and the basis for the technology planning. During the ideation sessions, stakeholders were interviewed to freely discuss potential scenarios independent of the available technologies and possible use cases. A maximum of 20 persons with different backgrounds and different profiles in ESA’s operations were asked on their ideas and on the use of this technology. Resulting scenarios have been grouped into main categories such as training, operations and planning. In a second step for each of the scenarios, representative use cases for the use of new technologies have been elaborated which lead to the identification of actors, their interactions, and the interaction with ESA’s ground systems resp. involved simulation modules from the ESA’s infrastructure (Fig. 1).

Fig. 1. Ideation and use case definition at ESOC/ESTEC/EAC premises (mid: high level scenarios and grouping, right use case definition describing (fltr) actors, tasks and interactions, simulation modules involved)

As a result, some of the scenarios within training should be mentioned here as those have been identified by several ESA units to be used for a potential proof of concept implementation. Here ideas grouped around: HW failure of a lunar rover and its recovery by an astronaut; HW failure in MCS Situation – problem occurs at MCS/fire or HW failure; training of personnel on recovery actions - A new set of astronauts arrive on lunar base, familiarizing themselves with existing infrastructure, inspecting the base, and astronaut has to assemble parts of VLB, e.g. new modules, antennas, parts of rover, etc. Training & familiarization of «any device» in VR assembly/disassembly of device in «micro or zero gravity» Moon Base operations – Astronaut performs maintenance task on lunar base: monitoring status of devices; performing actions; ground carries out "in parallel" involving bidirectional communication base and ground. The scenario of the astronaut-training center focused on an emergency fire case, in which astronauts have to be trained on how to deal with exceptional emergency operations, like rescue and evacuation, communication & coordination with ground, extinguish fire in module.

3.2 Selection of Technologies

VR

In a subsequent step, the choice of technologies has been taken in which recent technologies (a comprehensive overview of STAR technologies can be found in [5]) have been chosen in order to establish the VR infrastructure for the VLB. Here, following the taxonomy of [5], we have focused on the deployment of a stationary head mounted displays such as HTC Vive (see Fig. 2, top). For manual tasks we will evaluate ManusVR[1] gloves (see Fig. 2, right) as input device for gesture tracking and manual interactions as an alternative to the HTC Vive VR controllers (see Fig. 2, left). These

Fig. 2. VR set-up for the VLB implementation, top: HTC Vive HMD, left: Vive controllers, right: manusVR gloves.

[1] Manus VR gloves are wireless devices. They provide on each glove a sensor using IMU and bend sensor technology giving information about the position and orientation of the hand. manus-vr.com.

controllers are shipping with the HTC Vive and are based on their lighthouse tracking solution. For VR application development, we used Unity3D[2].

Simulation
While there are different simulation environments in use all over ESA, we identified SIMSAT as a good candidate for the PoC, because it is widely used and there are already interfaces for different other components relevant in different identified scenarios. In addition, this decision allowed us to use behavioral models made for the lunar base.

3.3 Realization

For the realization of the VR-DES coupling, we have defined an indirect connection of the VR environment to the SIMULUS/SIMSAT infrastructure of ESA. Here, the interface leverages a python adapter to connect via the CORBA API provided by SIMSAT. On the other side, the adapter connects to a message broker. This message broker is used to exchange messages between the SIMSAT adapter and the VR. Because its message protocol is open and the messages are JSON encoded, this interface can be used for other clients in the future. Possibilities are for example web based status displays or augmented reality systems, which are planned for a future study.

In the VR environment, a single node directly connects to the message broker, and handles the communication between special interactive nodes and the simulation. The system has been designed in a way that allows scene developers to set up a new VR scenario without any programming on the simulation connection side.

The first prototypical implementation has been realized and was defined within the framework of the EAC training program (see next chapter).

4 Implementation of a 'Virtual Lunar Base' PoC

For the final development of a proof-of-concept (PoC) showcase, we have chosen in agreement with several ESA stakeholders a scenario in which they deemed VR as most beneficial. Thus, together with EAC/ESTEC and ESOC, we have been further developing an astronaut training scenario with related use cases, which has been identified in stage one of the project.

4.1 Scenario

The developed scenario simulates an emergency in a lunar habitat in which an astronaut – the habitant – has to follow a standard procedure typically applied in orbital stations such as ISS. Here, a fire is detected in one of the racks inside the lunar base, and a habitant has to follow the procedures to extinguish it. These procedures are in place to minimize the risk for the responder as well as other habitants of the lunar base. Several objects to handle this task are available to the user. Some of these objects are connected to the

[2] Unity3D SDK (2018) – unity3d.com.

simulation, while others only exist in VR to allow a trainer to verify that certain steps in the procedure are taken. In order to keep the scenario evaluation friendly, some simplifications where made to allow untrained evaluation participants to solve the tasks as well. For example, a device used to measure combustion products in an enclosure simply shows a temperature. The following list contains some examples of objects, which can be found in the VR scenario. Some of these are invisible to the user, some are stationary and some can be carried around.

- *Fire Extinguisher*: A gas cylinder containing a finite amount of CO^2. It has a mountable nozzle to deliver the gas directly into a rack. The user can pick it up, mount the nozzle and operate it.
- *Fire Zone*: Invisible volume inside of a rack, which communicates with the simulation. The user can only interact with this zone via the extinguisher or the meter.
- *Thermometer*: The device is inspired by the CSA-CP used on the ISS. To simplify handling, the probing lance is directly connected to the device, rather than via a hose.
- *Lights and Buttons*: These directly communicate via messages with the simulation. There are environment lights, alarm lights and switches for light, ventilation and energy.
- *O^2 Mask*: The use of this item only allows the trainer to check if procedures are followed.
- *Torches*: These can be picked up and used to illuminate the room.

The O^2 mask and torches are examples for objects, which have no connection to the simulation at all. The mask is worn to signal the trainer that this self-protection step in the procedures is followed. The torches help the user to navigate the lunar base when the lights are switched of, but this interaction happens purely in VR.

The trainer is able to modify simulation parameters and can check how his trainees do react in the presented scenario. Possible modifications for the trainer include:

- Selection between multiple fire zones,
- Automatic fire alarm or user notification via “Green Card” (message like “you smell a burning odor”),
- Failed automatic deactivation of the inter module ventilation,
- Severity of the fire, which determines if it can be stopped by removing power from the rack or an extinguisher has to be used.

The evaluation section contains a description of a use case, which has been chosen to test how fast a new user can handle the given tasks.

4.2 Interaction Design and Multi User Set-Up

To help new users to understand how to interact in this virtual environment, we included interactive interaction courses for controllers and gloves. Figure 3 on the left shows such interactive instructions for the gloves. These courses lead the users through all possible forms of interaction and show which buttons/gestures can be used.

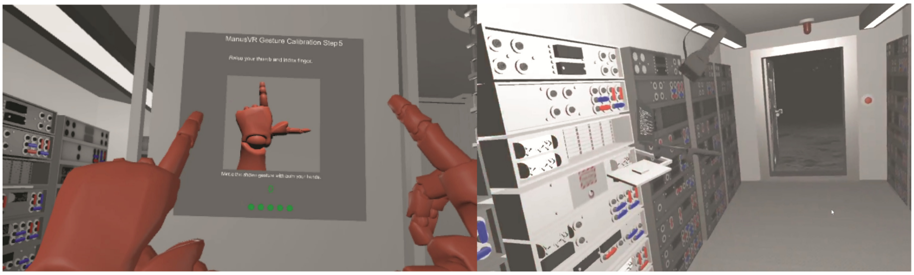


Fig. 3. Trainee perspective. Left: Calibration and interaction tutorial. Right: Sharing the space with a second trainee, who is represented by his HMD and controllers

Gloves and Controllers

One goal of the PoC was to have a platform, on which different interaction devices could be compared. During the workshop, we identified the Manus VR gloves as a more natural interaction device for VR, so the PoC was develop to be usable with both, Vive controllers and Manus VR gloves.

Trainer View

To observe what is happening in VR, the trainer can choose between different view modes: First Person, Overview and Follow. For First Person, the trainer shares the same camera perspective as the user. For Overview and Follow, the trainer gets a view from above the lunar base, with all structure above the users floor removed (see Fig. 4, right). For Overview, this view contains the whole base, while for Follow, the view is zoomed in on the user. When working with multiple Users, these options are available for every user.

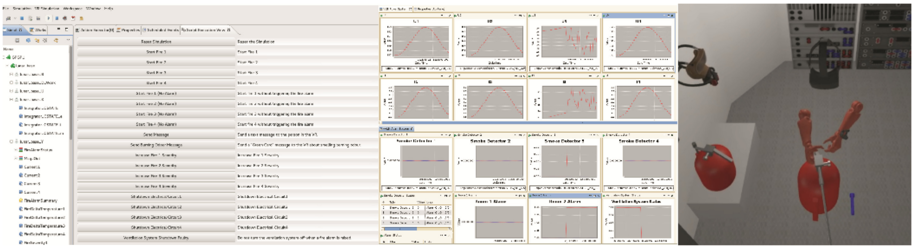


Fig. 4. Trainer Views. Left to right: Simulation interface, ground control telemetry and VR spectator view.

The trainer can also manipulate the simulation parameters to put his trainees in different situations. The left and center images in Fig. 4. Trainer Views. Left to right: Simulation interface, ground control telemetry and VR spectator view. Figure 4 show shortcuts to trigger common behavior in the simulation and a telemetry screen. He can also directly write into simulation parameters to create unique situations.

Multiple Users in VR

The tasks in the training session can be shared between multiple users. To allow this happens at two different levels: Simulation and VR. For the direct VR to VR connection, the capabilities of the VR-Engine where used. This allows the user to see each other in VR (represented by their interaction devices: HMD and Controllers/Gloves, see Fig. 3), and to synchronize the position of objects that can be picked up.

Object states that are controlled by the simulation do not need to be synchronized externally. Both VR instances react on the same messages, thus their objects have the same state. Therefore, if one user uses a light switch, his VR instance sends this message to the simulation and both VR instances receive the change of the lamp state.

5 User Studies and Evaluation

5.1 Methodology

In order to validate our solutions we conducted specific validation sessions with different ESA stakeholders addressing the fire emergency scenario in a simulated VR environment of the virtual lunar base. The system we were setting up offered the use of controllers for interaction, and, as it was required by the PoC, the use of gloves for an interactive manipulation of objects. Therefore, several validation sessions were conducted in two iterations. The simulation backend was controlled by a trainer that supervised the trainee during each session.

We decided on an easy to explain use case, which was tested by volunteers from the ESA staff (mostly ESOC and EAC members). Since the test users (trainee) came from different backgrounds and weren't necessary involved in the virtual lunar base project we constructed a set of tasks that were easy to explain without getting into too much detail about the environment.

To prepare the participants, they were given time to familiarize with the controllers while the trainer explained the function of different objects necessary for the tasks ahead. After that, the participants solved the tasks, while the trainer took time and gave verbal support if needed.

After this stage, the participant proceeded to repeat the experience with the gloves as input device. Since the setup for the gloves is significantly more difficult than for the controllers, we also measured the time it took the users to step through the glove calibration process.

An interactive guide then explained the participant which gestures are used to trigger the interactions learned with the controller, followed by a moment of time to test these in the environment. After this familiarization phase, the known tasks were repeated using the gloves.

Finally, the users had to fill in a questionnaire following the design mentioned below. We then could gather the feedback remotely and could analyze the users reaction to the developed PoC.

User Experience (UX)

We aligned our feedback design to recent research results provided by the HCI community. However, in contrast to only focus on pure usability aspects of the PoC, we aimed at analyzing the user experience (UX). In contrast to usability procedures, UX tries to balance instrumental and non-instrumental qualities that are subjectively perceived by the user. Thüring and Mahlke [6] claim, that the three central components of UX are represented by the perception of instrumental and non-instrumental qualities that trigger emotional reactions when a person uses or has used a product. They highlight that traditional usability testing with a focus on removing interaction problems, which cause stress, does not interfere enthusiasm for a product compulsory. Therefore, not only pragmatic aspects such as utility or usability of a product but also hedonic aspects such as stimulation, identification or evocation might influence the user perception. Those can then be measured by emotional reactions such as pleasure, satisfaction and develops an appeal towards the product.

Thus, the model of [6] guided the design of our questionnaire and the choice of specific aspects for "quantizing" user experience aspects. Due to the manifold of (non-)instrumental qualities, the selection has been based on "subjective" priorities for the purpose of this study. We have been focussing on three selected aspects of UX proposed by [7]:

- *perspicuity* as a measure of learnability – "How easy to understand is the user interface?"
- *efficiency* as a measure to what degree the user can use the interface with a high level of productivity – "Is the workload for the interaction with the VR environment reduced to a minimum?"
- *stimulation* as measure for the stimulus of a user, as a "driving power" for interaction and learning of new skills – "Does the VR environment captivate the user?"

Questionnaire

An online questionnaire suited as basis to gather feedback from the test subjects. The intention was to collect qualitative as well as quantitative data during each validation session. The time required to conduct certain tasks following the standard procedure is considered as one of the quantitative parameters within the subsequent evaluation for the execution with controllers and gloves, the elevation of the non-instrumental and instrumental aspects for perspicuity, efficiency, and stimulation the others. Here, we mapped the aspects to a Likert scale with 7 features ranging from one to another extreme. For a later assessment, the scale was subdivided into five intervals that are allocated to different levels of maturity [8]. The neutral interval contains values between [−0.8; 0.8 (To achieve a "good" the interval has been defined to [0.8; 2] resp. "poor" (−0.8; −2]. Extreme values at the end of the scale are rated as either "very poor" or "very good" (Fig. 5).

Interval	[-3,-2[	[-2,-0.8[	[-0.8,0.8]	]0.8,2]	]2,3]
Label	very poor	poor	neutral	good	very good
Notation	--	-	o	+	++

Factor	Extract from the User Experience Questionnaire	1	2	3	4	5	6	7		
Perspicuity	not understandable	○	○	○	○	○	○	○	understandable	1
Efficiency	organized	○	○	○	○	○	○	○	cluttered	2
Stimulation	motivating	○	○	○	○	○	○	○	demotivating	3
Efficiency	inefficient	○	○	○	○	○	○	○	efficient	4
Perspicuity	complicated	○	○	○	○	○	○	○	easy	5
Stimulation	boring	○	○	○	○	○	○	○	exciting	6
Perspicuity	clear	○	○	○	○	○	○	○	confusing	7
Efficiency	fast	○	○	○	○	○	○	○	slow	8
Stimulation	not interesting	○	○	○	○	○	○	○	interesting	9
Perspicuity	easy to learn	○	○	○	○	○	○	○	difficult to learn	10
Stimulation	valuable	○	○	○	○	○	○	○	inferior	11
Efficiency	impractical	○	○	○	○	○	○	○	practical	12

Fig. 5. Instrumental and non-instrumental aspects elevated in this study. Mapping them onto a Likert scale following a typical user experience questionnaire layout [8]

The qualitative data blocks contained five questions with pre-formulated answers following the layout in [7] and a free text area where subjects could comment. This data blocks should reflect on the user reactions supporting the capturing of certain UX aspects. We asked:

1. What do you think about the handling of the VR Scenario using the devices?
2. What do you think about the time it took to use the devices?
3. What do you think about your experience with the VR scenario?
4. What do you think about the functions offered by the VR environment and the presented scenario?
5. Was every function available that you wish to have?

5.2 Validation Sessions

Place and Time

The validation sessions were conducted during a two-week period at the premises of ESOC in Darmstadt and the European Astronaut Training Center (EAC) in Cologne.

Participants

The participants were volunteers from ESOC and EAC. Roughly ~2/3 were male. The youngest participant was 20 and the oldest 55, with a median at 30. About 10% claimed to be experienced VR users, while the rest divided equally into users with some and with no experience. In total, we had received 17 filled in questionnaires.

Test Cases

The volunteers had to fulfil the following tasks that have been presented at the initial stage of the validation session and the methodology described above. Figure 6 shows some of these tasks:

1. You will be alerted by the trainer about high temperature readings in a specific rack.

2. Take the temperature probe and verify the high temperatures in the specified rack.
3. Switch off the power circuit for the rack.
4. Pick up the oxygen mask and mount it onto your head.
5. Pick up the fire extinguisher and mount the barrel to the valve Sect.
6. Empty the extinguisher into the rack.
7. Use the Temperature probe to verify that the fire is extinguished and alert the trainer.

Fig. 6. Evaluation Tasks – Left to Right: checking temperature, switching off power, putting on the oxygen mask and extinguishing the fire

5.3 Results

Quantitative Analysis

We have gathered the feedback from the Likert scale and normed then according to the described procedure. Interesting wise, it clearly shows high values for the controllers. Although the median values are not too far from each other, the variance for using glove control is much higher, to the upper 75% percentile as well as to the lower 25% percentile in several aspects. Especially the lower whiskers, which range to "bad" indicate that the users were not very satisfied using glove control. The perspicuity of glove control has been perceived as much more complicated than using the controllers. Also, users have not been convinced to execute the tasks in an efficient way as compared to the execution of the tasks using the controllers (Fig. 7).

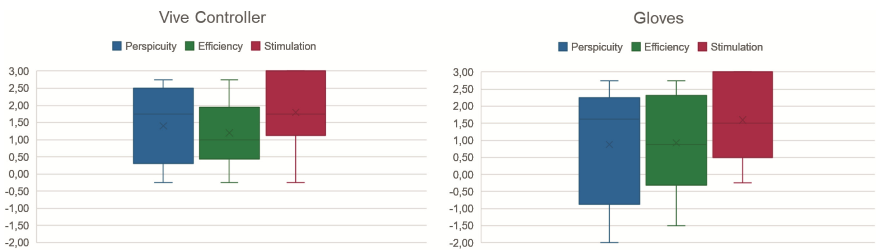


Fig. 7. Instrumental and non-instrumental aspects for controller and gloves

The summary chart Fig. 8 visualizes the aggregated mean values over the Likert scale results for each aspect.

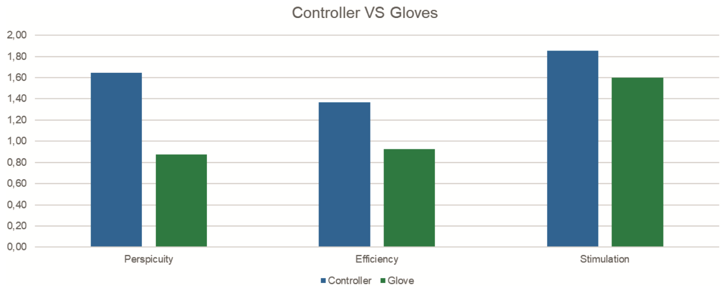


Fig. 8. Comparison of the averaged results for perspicuity, efficiency and stimulation with controllers and gloves

Timing of the Tasks

We have captured the timing for execution of the tasks in both system settings. In both settings, the trainer defined a fire location in order to start the simulation. After the trainee indicated a "ready" to the trainer, the trainer started the time recording. It was stopped as soon as the tasks have been fully executed and the fire simulation has been extinguished resp. the temperature went below a lower degree threshold. Figure 9 shows the results of the timing. In addition, here, the volunteers managed to solve the tasks faster using the controllers rather than the gloves. However, median values are not too distant; the discrepancy to the upper percentile (0.75) and upper whisker is more significant for the gloves. A clear timing advantage has been observed for the controllers.

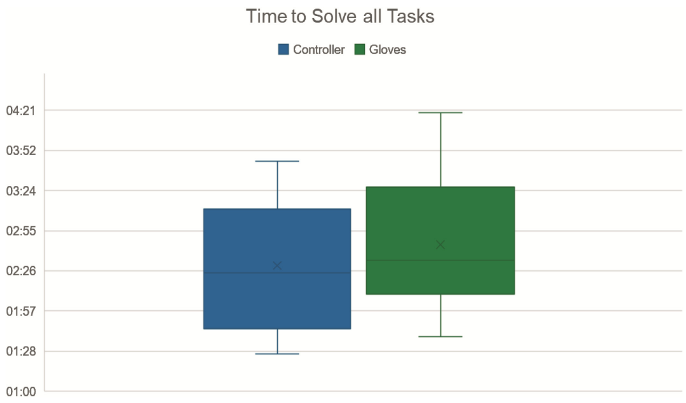


Fig. 9. Results of the time recording to solve all tasks

Conclusion

While all users had their experience with controllers first, we expected to see faster times with gloves, since the tasks are only a repetition. While the overall feedback is positive, the feedback indicates a more positive response to and a faster execution time using the controller interaction. Taking into account, an average calibration time of

1:52 min for glove control, the controller interaction became more accepted. This quantitative feedback is supported by the results of the qualitative data blocks.

Qualitative Analysis

Testing with the Controllers
While 94% rated the experience as completely useful and interesting, only 75% found it to have all functionality they wanted to have. While one user missed feature in the physical modelling, others wished to have the means to do moon walks or have simulated smoke in the habitat.

Only 6% deemed the handling of the VR to be too demanding to focus on the content, 41% needed some time before they were able to focus and over 50% could had no problems to use it from the start.

82% of participants were satisfied with the time it took them to solve the tasks, and the remaining 18% had solvable problems to some extent. None of the participants rated these problems high enough to say they prevented them from consuming the content.

From the free text feedback, most of the comments referred to aspects of the virtual lunar base that were excluded from the evaluation scenario. Others made clear that we are missing a clear optical distinction between indicators and switches, likewise, that there is no way for the user to know if he is hitting the fire zone with his extinguisher nozzle. Another request was for a more game like interaction, where objects are collected and can be directly at hand if needed, instead of carrying every object on its own. One participant asked for direct movement instead of teleportation, which let us know that when explaining teleportation, we should tell how it helps to prevent motion sickness in many users.

Testing with Gloves
30% of users felt that the glove interaction was too complicated to focus on the content, and another 60% needed some time before they could focus. Only a third was satisfied with the time it took them to solve the tasks.

In the free text feedback, some participants showed that they enjoyed the more natural interaction, once they familiarized themselves enough with the gloves. Other where astonished that the gloves were not easier to use than the controllers were. Some users complained that not all interactions using the gloves are "natural", which is to be expected for teleportation, but also grabbing an object can hardly be experienced in a natural way without haptic feedback.

6 Conclusion

This paper presented the results of a one-year study in order to elevate the potential of ESA's space operations being linked to a simulated training environment. We have realized a virtual lunar base that aimed at establishing a training environment for "lunar exploration", which has been identified by ESA as one likely next step of human space exploration beyond the ISS. We managed to couple the DES system backend of ESA's simulation environment to a highly interactive virtual environment, in which astronauts can be trained for certain tasks that a trainer could design and develop. The exemplified

PoC realization focused on a dedicated training scenario that has been developed together with ESA's astronaut trainers as well as further stakeholders at the different premises of ESA. The overall decision chain and communication flow between lunar base, mission control and the link to ESA's telemetry simulation backend has been realized. A fully-fledged fire emergency simulation can be performed in the VR environment, leveraging several physical simulation results as well as ESA's SIMSAT backend services. In order to validate the different possibilities for user interaction using VR controllers as well as recent glove solutions (here ManusVR), dedicated validation sessions have been set-up, the feedback of volunteers gathered and analyzed. Our initial hypothesis that the comfortable ManusVR gloves might be more attractive to the volunteers and might lead to a better task performance over VR controller could not be kept. Instead, the overall performance of volunteers dropped behind in several instrumental and non-instrumental aspects of user perception compared to VR controllers. More frustratingly, the timings the volunteers needed to execute a range of tasks using classical controllers are lower than with gloves. Adding the calibration time that each user needed for using the gloves to even get started, only a weak recommendation for a similar system set-up with gloves can be given.

References

1. ESA call for tender, Statement of Work, Doc. Ref. ESA-DHSO-SYS-SOW-0002, published 30 May 2016
2. Akpan, J., Brooks, R.J.: Practitioners' perception of the impacts of virtual reality on discrete-event simulation. In: Proceedings of the Winter Simulation Conference, pp. 1976–1984 (2005)
3. Pulugurtha, S., Nambisan, S., Dangeti, M., Kaseko, M.: Simulating and analyzing incidents using CORSIM and VISSIM traffic simulation software. In: 7th International Conference on: Applications of Advanced Technology in Transportation (2002)
4. Turner, C.J., Hutabarat, W., Oyekan, J., Tiwari, A.: Discrete event simulation and virtual reality use in industry: new opportunities and future trends. IEEE Trans. Hum.-Mach. Syst. **46**(6), 882–894 (2016)
5. Anthes, C., García-Hernández, R.J., Wiedemann, M., Kranzlmüller, D.: State of the art of virtual reality technology. In: Proceedings of the IEEE Aerospace Conference, Big Sky, MT, USA (2016)
6. Thüring, M., Mahlke, S.: Usability, aesthetics, and emotion in human-technology interaction. Int. J. Psychol. **42**, 253–264 (2007)
7. Gockel, B., Graf, H., Pagano, A., Pescarin, S., Eriksson, J.: VMUXE: an approach to user experience evaluation for virtual museums. In: Marcus, A. (ed.) DUXU 2013. LNCS, vol. 8012, pp. 262–272. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39229-0_29
8. UEQ Online, here downloadable Evaluation Kit, English Version. http://www.ueq-online.org/?page_id=8. Accessed 21 Jan 2013

Guiding or Exploring? Finding the Right Way to Teach Students Structural Analysis with Augmented Reality

Rafael Radkowski[1(✉)], Aliye Karabulut-Ilgu[2], Yelda Turkan[3], Amir Behzadan[4], and An Chen[2]

[1] Virtual Reality Applications Center, Iowa State University, Ames, IA, USA
rafael@iastate.edu
[2] Civil Engineering, Iowa State University, Ames, IA, USA
[3] Civil Engineering, Oregon State University, Corvallis, OR, USA
[4] Civil Engineering, Texas A&M University, College Station, TX, USA

Abstract. The paper reports on the design of an augmented reality (AR) application for structural analysis education. Structural analysis is a significant course in every civil engineering program. The course focuses on load and stress distributions in buildings, bridges, and other structures. Students learn about graphical and mathematical models that embody structures as well as to utilize those models to determine the safety of a structure. An often reported obstacle is the missing link between these graphical models and a real building. Students often do not see the connection, which hinders them to utilize the models correctly. We designed an AR application that superimposes real buildings with graphical widgets of structural elements to help students establishing this link. The focus of this study is on application design, especially on the question whether students prefer an application that guides them when solving an engineering problem or whether the students prefer to explore. Students were asked to solve a problem with the application, which either instructed them step-by-step or allowed the students to use all feature on their own (exploring). The results are inconclusive, however, tend to favor the explore mode.

Keywords: Augmented reality · Education · Engineering · Structural analysis Virtual instructions

1 Introduction

The goal of our research is to help students to better understand structural analysis using Augmented reality (AR). Structural analysis is a significant course in any engineering program. Students learn about load and stress distributions in buildings, bridges, and other structural elements. The course program incorporates the structural models such as beams, trusses, and others, as well as their utilization to determine the safety of a structure. All structural models, also called members, usually have a graphical component and a mathematical description. The first one allows students and engineers to sketch the load elements of a structure, the second one allows for analyzing the structure's load. Figure 1 shows an example of a frame structure. Students learn in a course

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 452–461, 2018.
https://doi.org/10.1007/978-3-319-91581-4_34

how they can use it theoretically. However, to connect it to a building such as the one Fig. 1 illustrates remains challenging.

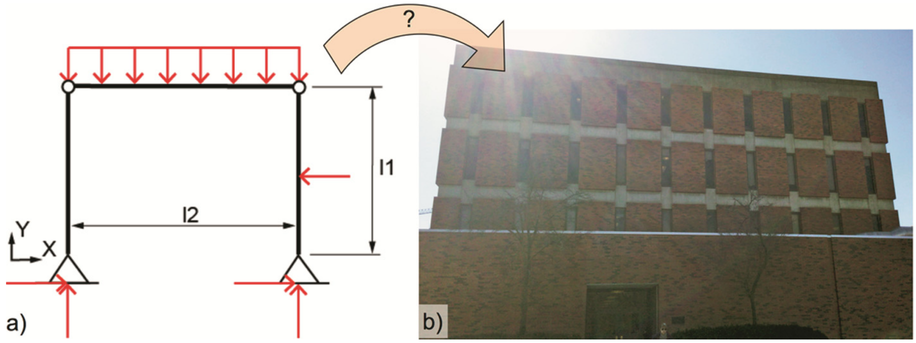


Fig. 1. (a) A typical frame structure engineers use to analyze the load and deformation of buildings. (b) The frame can represent parts of the building. Students often struggle to see the relation.

The struggle for novice learners is twofold: First, it is difficult for students to identify the correct member for a particular building (i.e., beam, truss, frame). Second, even if they found the element, it is often not clear how one has to represent a real building using the graphical and mathematical components. More massive structures incorporate multiple connected elements, connected by supports and other means. As a consequence, most novices in this course do not appear to gain the required understanding of fundamental concepts, such as load effects and load paths. They also cannot visualize the deformed shape of simple structures, a necessary skill to conceptualize structural behavior beyond formulas and methods [1, 2]. The prevailing theory attributes those learning deficiencies to the ineffectiveness of the traditional lecture-mode of teaching. Instructors spend much effort on the analysis of discrete members, and less on understanding the behavior of the entire structure in a three-dimensional context.

We developed an AR application denoted as iStructAR to supplement traditional structural analysis lectures. The application superimposes a building or any other structure with graphical widgets indicating members, supports, loads, and other elements. The application is interactive. Students can change, for instance, the magnitude of a load, its distribution, and other required elements to solve an engineering problem. We initially introduced the application in [3, 4]. This research focuses on the mode of instruction: Guiding or Exploring. Students of an introductory structural analysis course are novices to this topic. Consequently, an instructor guides the students through the problem-solving process before he or she encourages them to solve a problem on their own accord. Following this notion, the application may also guide the students. On the flipside, it may just act as a facilitator that encourage the students, nevertheless, requires them to know the problem-solving process. iStructAR implements the second option. However, it encountered opposition when students of an initial study [1] reported they "I did not know what to do." To further assess these options, we realized a second, guided application mode, which takes the students through a step-by-step procedure. Students that participated in a study had to either use the guided mode or the exploring mode.

The next section introduces the application. Section 3 explains the study and reports the results. The last section closes with a conclusion and future aspects of this research.

2 iStructAR - Application Design and Scenarios

iStructAR is an application designed for an Apple iPad. Its interface and entire design utilize the standard Apple UI widgets (Sect. 2.1). To support typical lectures, it implements a set of structural analysis scenarios that supplement the problems instructors assign to students (Sect. 2.2). Each scenario addresses a different aspect of structural analysis.

2.1 Application Design

Figure 2 shows a screenshot of one of the application scenarios. The application operates in landscape model and shows a video of a selected building. Here, it shows a skywalk, the structure of interest. The graphical interface follows a typical border layout with the center area in the center surrounded by interaction widgets along the border. Since the application utilizes AR, the central area displays a video captured by the iPad's back camera, which the renderer superimposes with 3D models representing different structural members. Several graphical widgets along the border allow a user to control the application. All widgets are organized by functions groups. The bar at the bottom contains widgets that allow the user to control the app. These widgets include, for instance, button icons to change the load situation. Since the load situation changes for different scenarios, the definition of these buttons can change. The right side of the bottom bar hosts control button icons. A user can control the usage mode (stand along, indoor, outdoors). At the upper left corner, the user can find several switches to turn on or off visualization widgets. These switches allow a user to enable or disable particular load visualizations, independently from the load itself. Depending on the task, the display can become cluttered with graphical widgets. Thus, a student might lose track of all the different items on the screen. Being able to switch specific features on and off helps to maintain the overview. Besides using button icons and switches to interact, all loads widgets on the primary display are also gesture controlled. The user can change the location and the magnitude of a load with his/her fingers.

From an application point-of-view, students should work with the application in front of a physical building[1]. The application works with object recognition. Thus, it recognizes when a student points the iPad towards, e.g., the skywalk, it calculates the pose, and instantly populates the image with 3D models indicating the load, reaction forces, and other scenario-dependent items. It is possible for a student to freeze the video feed

[1] A video demonstrating the usage can be found on YouTube: https://www.youtube.com/watch?v=kANM-r5Fh0k.

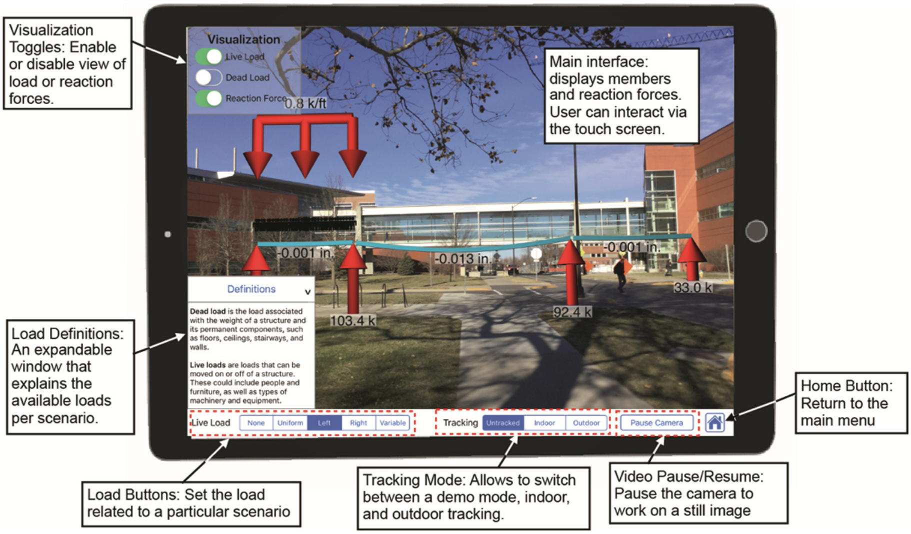


Fig. 2. The interface structure of iStructAR.

and work with the last image instead. Pointing the iPad towards a building is tiring, so this option allows a student to work in a more comfortable posture.

The application also supports an indoor mode. During winter times or severe weather conditions, students can work with printouts of the structure. Although this solution is not ideal since they see the building from one perspective only, it is an acceptable tradeoff between good-weather and not using the application at all. The last mode is a so-called demo mode. This mode only shows a prepared image of the building. It acts as a demonstrations scenario to quickly show the application features. Students should not use it.

2.2 Application Scenarios

Currently, the application content is still under development. However, students can already practice their skills on three engineering application scenarios. Each application scenario is either related to a particular structural member (i.e., beam, frame, truss) or a load situation (i.e., static load distributions, wind load, seismic load).

Figure 3 depicts the beam application scenario. The objective of this scenario is to explain beam deflection under load. Every solid surface deflects when one exerts load to it. In this scenario, the load comes in the form of a so-called dead load, which is due to the weight of the structure, and a live load, due to the people crossing the skywalk.

Students need to understand how different load situations affect the deflection, especially the form of the deflected beam. For that reason, an instructor would usually assign a task to students that asks them to draw the deflected beam for different load situations and to compare the form of the beam. Instead of drawing, the application allows the students changing the load, its distribution and magnitude in particular, and to observe the differences. The essential point here is that students can instantly relate the form of the deflection to the concrete skywalk and notice the critical locations. When comparing

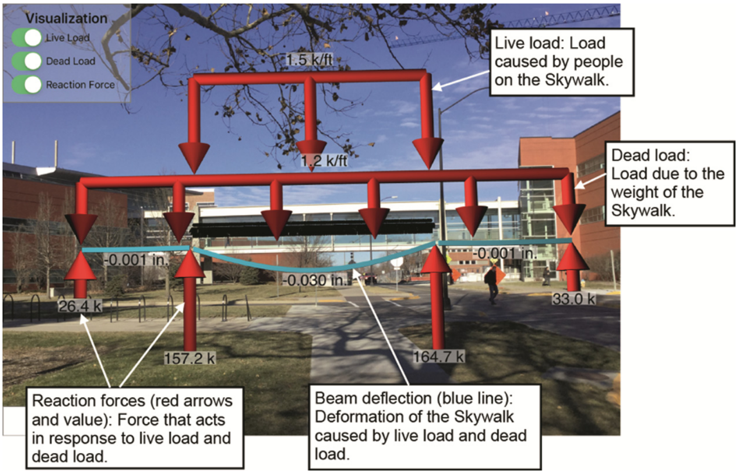


Fig. 3. The Skywalk scenario features a beam (blue line) Students can change the load, the application calculates and visualizes the response (deflection and reaction forces). (Color figure online)

multiple load situations, students should be able to develop an understanding which sections of a skywalk are the most critical once and how a beam member allows them to identify these critical points.

Figure 4 demonstrates a seismic load scenario. Earthquakes exert a seismic load on buildings. It appears as an oscillating load, which shakes the building. The challenge here comes in the form of the technical definition of the seismic load. Civil engineers employ a seismic map which states characteristic factors for every region of a country. Although these characteristic factors encode the load, engineers need to translate them it into an oscillating load first. Thus, those factors have no natural meaning. The AR application visualizes the load and the deflection. Load distribution and deflection change quadratic and can exceed a critical tolerance at a certain point. In this example, using a bell tower, the particular point is a certain height. Using the AR applications, students can observe the tower from multiple directions and notice the critical heights. Furthermore, they can change the characteristic factors. Changing factors technically means they virtually place the bell tower in a different location. Again, they can compare the different critical loads and deflections and analyze whether a bell tower build for Iowa would also last in California, for instance.

The hypothesis is that the AR application facilitates learning because the students can better associate a structural member or a response to a load to a building. Furthermore, the application should engage students to work on the topic. Therefore, the application design follows three principles, which were often favored in the related studies [5–7]: The content is semantically context-related, interactive, and provides instant feedback. Registered 3D models of members and load realize semantically context-related content. We hypothesize that this will allow students to better understand how they can utilize structural members. Instant feedback to interactive changes should also

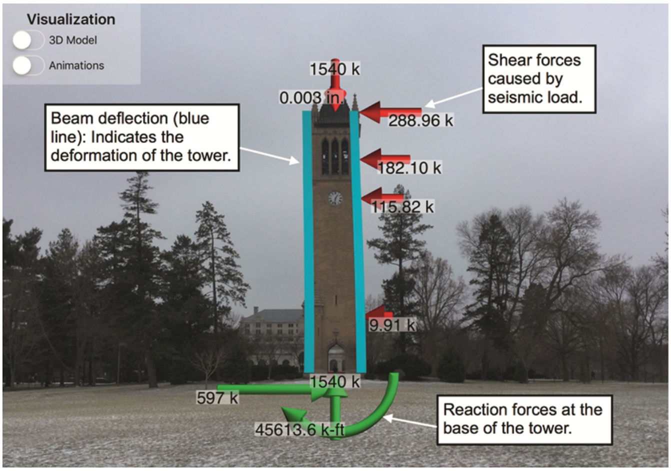


Fig. 4. Seismic load on a tower. This scenario shows how a tower is modeled with beam members. The scenario shows the reaction forces and deflection due to seismic stimulation.

allow them to better learn the effects of load changed. Students can experiment with load situations and compare the results on structures.

3 Study

We encountered a challenge pertaining application usage during our initial study [1]. Students reported they were unsure how they can use the application to solve a problem the instructor assigned to them. With this result faced, we decided to test whether students would prefer a more guided approach. One that tells or asks them step-by-step what to do. Although this might betray the purpose of the entire class, students of this course are still novices and might be overwhelmed when already left alone with a problem after one first example. In an ideal case, students should experiment and explore until they can solve problems on their own accord. To better understand the struggle, we prepared two applications modes denoted as guided-mode and explore-mode (Sect. 3.2). Students were asked to solve a problem (Sect. 3.1) using one of these modes.

3.1 Problem

The problem incorporates three questions that the students had to solve using the application.

Question 1: "If a uniformly distributed load was placed over the structure shown below, how would each span deflect? Draw the deflection shape on the diagram below" (Fig. 5a).

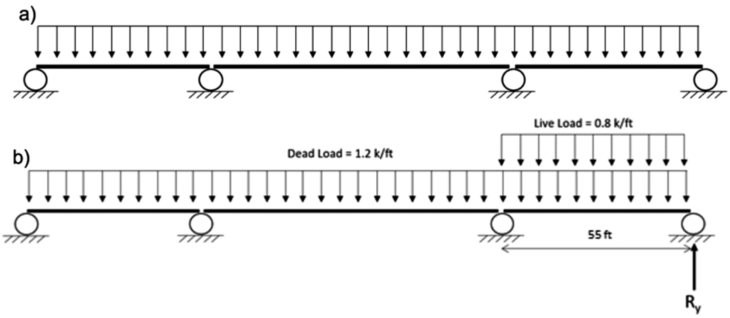


Fig. 5. a–b: Supplementing figures that visualize the load situation of the problems.

Question 2: "If students were to walk across the entire Skywalk, or stand on the entire Skywalk, what type of load could be used to model their effect on the structure? And, how would this load affect the deflection? Draw the deflection from the additional load on your original deflection image."

Question 3: "The diagram below has the same span lengths as the modeled Skywalk in the app. The third span length is 55 feet. If a uniformly distributed dead load of 1.2 k/ft was on the entire structure, and a uniformly distributed live load of 0.8 k/ft was on the third span, what would the reaction at the fourth support (shown as Ry) be? Remember that each individual span is simply supported." (Figure 5b) shows the supplementing sketch

A student can solve all three questions with the application by adjusting application parameters until they match the stated situation. The first question requires the student to adjust the load on a beam. Hence, the student needs to identify a beam (Skywalk scenario) and adjust the load. As a result, the graphical beam would deflect and demonstrate the deflection. The procedure to solve the second question does not deviate much from the first one. Here, a student is required to know how he or she can model group of scattered people on a Skywalk. A solution for that is also part of the application scenario. Likewise, the last question asks the student to set the load parameters correctly.

Each question imposes two problem-solving aspects to a student. He or she first needs a minimum understanding of structural analysis. This aspect pertains structural analysis. Secondly, he or she also has to identify the graphical widgets allow one to change load parameters and visualization settings, which is a usability aspect.

3.2 Guided-Mode Vs. Explore-Mode

We prepared two versions of the application denoted as guided-mode and explore-mode.

The guided-mode provides step-by-step instructions as text to a student (Fig. 6). All instructions appear on the lower bottom of the screen. A "next" button icon allows the student to move through all instructions for this task. With "previous" they can step back. Currently, the application does not verify whether a student completed a step correctly. A student can also browse through all instructions first before attempting to solve the problem.

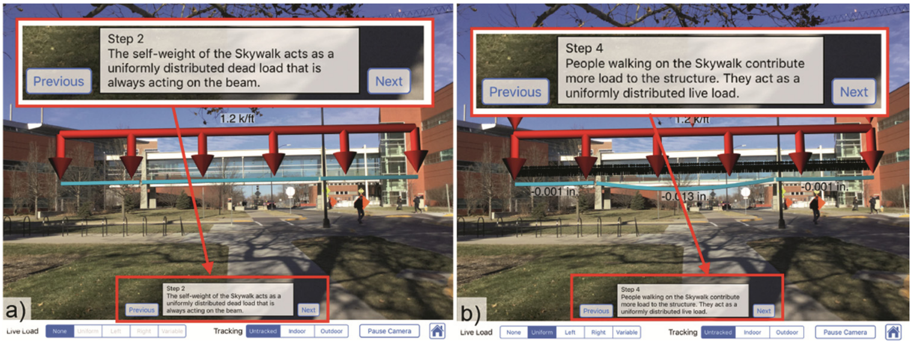


Fig. 6. a–b: A window at the bottom of the screen conveys instructions.

The explore-mode allows the students to use the application as it is; as described before. We did not add any additional hints. This mode requires the student to possess some problem-solving skills or at least, need the student to be explorative so that he or she can understand how s/he can use the application to solve the problems.

On a difficulty scale, we consider the three questions as simple, on the right level for novices who just start to learn about structural analysis.

3.3 Experiments and Results

We recruited students from the civil engineering department to conduct experiments following a between-subject design. Upon arrival, each volunteer read and sign an informed consent document. Next, an experimenter introduced the tasks, the equipment and the procedure to the student. Next, every volunteer completed a brief prequestionnaire which record age, sex, level of education, and field of study or profession. The student could ask questions. The experiment started when the student had no further questions. There was no time limit imposed, although the experimenter asked the student to be as fast as possible. Each student completed a questionnaire afterward. The questionnaire asked 16 questions (Table 1). We used a 5-point Likert scale. Students were also asked to take screenshots.

Table 1 shows the results of the questionnaire. We calculated the mean and standard deviation for each question, split into the guided mode and explore model. The overall mean for the guided mode is $\mu_g = 4.08$ (std. dev = 0.62) and for the explore mode $\mu_e =$ 4.40 (std. dev = 0.67).

The results are inconclusive. However, students tend to favor the unguided mode. Although the mean values for all questions do not yield any significant results, the variance indicates a particular trend towards the unguided mode. For unknown reasons, the students that used the guided mode tends towards problems finding the right buttons and graphical widgets that they need to solve the problem. The questions 1, 4, 6 and 7 indicate this trend given the variance within all answers. In general, all results indicate that the students only had little problems to use the application to solve the given problems. We initially assumed that students might have problems identifying beams or may

Table 1. Results of the post questionnaire.

No	Question	Guided mode		Explore mode	
		Mean	Std. dev	Mean	Std. dev
1	I could identify the functionality of the available options through icons	4.00	1.00	4.33	0.58
2	I could select the desired functions on the app all the time	4.00	0.00	4.67	0.58
3	I could easily undo/redo any action if I felt to do it	4.67	0.58	5.00	0.00
4	I could easily figure out what to do next given the instructions from the test facilitator	4.00	1.00	5.00	0.00
5	I could understand the messages that appeared in the app	4.67	0.58	5.00	0.00
6	I was not confused or lost while performing the tasks	3.67	0.58	3.67	0.58
7	I could enable the visualization of the load	4.33	0.58	5.00	0.00
8	I could easily distinguish the different types of load on the screen	4.67	0.58	4.67	0.58
9	I believe the arrows are appropriate representations of loads on the structure	4.67	0.58	5.00	0.00
10	I believe the curved line (the blue color line) is a good representation of the structure's deflection	4.33	0.58	4.67	0.58
11	I could easily read the distributed load values, reaction force values, and deflection values on the screen	4.67	0.58	5.00	0.00
12	I could easily take a screenshot of the app screen	3.33	0.58	4.00	1.00
13	I could easily change the load to a certain magnitude given by the test facilitator	3.00	1.00	3.33	2.08
14	I could easily manipulate the location of the distributed load to a certain position	4.33	0.58	3.67	1.53
15	When I manipulate the location of the distributed load, I prefer dragging the load on the touch screen over clicking button icons	2.33	0.58	3.00	2.00
16	The definitions of dead load and live load contained within the app interface helped me to understand the different types of loads	4.67	0.58	4.33	1.15

struggle to distinguish the different types of load. The results to question 9 and 10 indicate the opposite, regardless of which mode a student worked with. Despite the fact that this result is also not significant, the overall trend also shows that the interface design complies with the students' expectations for graphical representations of structural analysis members. A surprise to us was the students' problem to take screenshots using an iPad; they were asked to take a screenshot to document their results. The application interface had no button to save screenshots at this time and we assumed that students know how they have to operate an iPad to record screenshots. Apparently, this assumption was wrong.

A limitation of this study is currently the number of scenarios that were tested. Students had to solve a beam problem only. Thus, other members such as frames and trusses remain untested. Although we currently see no essential differences from a problem-solving perspective, no data is available to underpin this assumption.

4 Conclusion and Outlook

Due to the insignificant results, we cannot state any certain conclusion. However, the initial plan for this project intends to use an explore mode, which allows students to gain a better understanding by exploring a solution on their own accord. The results at hand do not indicate any difference. Thus, the authors will maintain the explore mode solution for future studies. Additionally, and despite our initial doubts after testing, the authors also do believe that allowing students to explore a solution to a problem is a better teaching strategy. Guiding students is almost comparable to giving them a solution. Almost, since the students still have to connect the dots, e.g., identify a truss or beam and the effects of load on those structures on their own. Nonetheless, it denies them the experience to determine the procedure on their own accord.

A next step is a study in a classroom with two sections of a structural engineering class at Iowa State University. One section, the students, in particular, will pose the experimental group working with the iPad, the other the control group. We scheduled the study to be conducted during the Fall 2018 term.

Please visit web.me.iastate.edu/istructar for future update.

Acknowledgment. This project is funded by the National Science Foundation under grant number #1712049.

References

1. Davalos, J.F., Moran, C.J., Kodkani, S.S.: Neoclassical active learning approach for structural analysis. In: Proceedings of the 2003 American Society for Engineering Education Annual Conference & Exposition (2003)
2. Teng, J.G., Song, C.Y., Yuan, X.F.: Fostering creativity in students in the teaching of structural analysis. Int. J. Eng. Educ. **20**(1), 96–102 (2004)
3. Turkan, Y., Radkowski, R., Karabulut-Ilgu, A., Behzadan, A.H., Chen, A.: Mobile augmented reality for teaching structural analysis. Adv. Eng. Inform. J. **34**, 90–100 (2017)
4. Turkan, Y., Chen, A., Karabulut-Ilgu, A., Radkowski, R., Jahren, C., Behzadan, A.H.: Mobile augmented reality implementation for structural analysis. In: 23rd International Workshop of the European Group for Intelligent Computing in Engineering (EG-ICE), Krakow, Poland, June 2016
5. Singhal, S., Bagga, S., Goyal, P., Saxena, V.: Augmented chemistry: interactive education system. Int. J. Comput. Appl. **49**(15), 1–5 (2012)
6. Qassem, L.M., Hawai, H.A., Shehhi, S.A., Zemerly, M.J., Ng, J.W.: AIR-EDUTECH: augmented immersive reality (AIR) technology for high school chemistry education. In: 2016 IEEE Global Engineering Education Conference (EDUCON), pp. 842–847 (2016)
7. Li, N., Gu, Y.X., Chang, L., Duh, H.B.L.: Influences of AR-supported simulation on learning effectiveness in face-to-face collaborative learning for physics. In: 2011 11th IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 320–322 (2011)

Assembly Training: Comparing the Effects of Head-Mounted Displays and Face-to-Face Training

Stefan Werrlich[1(✉)], Carolin Lorber[1(✉)], Phuc-Anh Nguyen[1], Carlos Emilio Franco Yanez[2], and Gunther Notni[3]

[1] BMW Group, Petuelring 130, 80788 Munich, Germany
{stefan.werrlich, carolin.lober}@bmw.de
[2] ITESM, San Luis Potosí, Mexico
[3] Group for Quality Assurance and Industrial Image Processing, Technical University Ilmenau, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, Germany

Abstract. Due to increasing complexity of assembly tasks at manual workplaces, intensive training of new employees is absolutely essential to ensure high process and product quality. Interactive assistive systems are becoming more and more important as they can support workers during manual procedural tasks. New assistive technologies such as Augmented Reality (AR) are introduced to the industrial domain, especially in the automotive industry. AR allows for enriching our real world with additional virtual information. We are observing a trend in using head-mounted displays (HMDs) in order to support new employees during assembly training tasks. This technology claims to improve the efficiency and quality of assembly and maintenance tasks but so far, HMDs have not been scientifically compared against face-to-face training. In this paper, we aim to close this gap in research by comparing HMD instructions to face-to-face training using a real-life engine assembly task. We executed a training-session with a total of 36 participants. Results showed that trainees who performed the assembly training with HMD support made 10% less picking mistakes, 5% less assembly mistakes and 60% caused less rework but they are significantly slower compared to face-to-face training. We further aimed to rate user satisfaction by using the system usability scale (SUS) questionnaire. Results indicated an average SUS of 73,5 which means ‘good’. These and further findings are presented in this paper.

Keywords: Augmented Reality · Evaluation · Head-mounted displays Training

1 Introduction

The automotive industry is undergoing radical changes due to growing customer demands and alternative business models. Increasing product individualization and diversity generates a transformation of the classic manufacturing environment. Modern production lines are built in a lean, flexible and adaptable way with a high level of automation in order to increase productivity. These challenges emphasize the

J. Y. C. Chen and G. Fragomeni (Eds.): VAMR 2018, LNCS 10909, pp. 462–476, 2018.
https://doi.org/10.1007/978-3-319-91581-4_35

importance of well-educated employees [1]. Training of novice shop-floor operators is essential when complexity in future production systems increases. During traditional face-to-face training, several assistance systems such as pick-to-light and information screens are already being used. Furthermore, new promising technologies such as Augmented Reality (AR) are introduced slowly to the industrial domain. AR allows to superimpose a realistic environment with additional virtual information. Hand-held devices, projectors and head-mounted displays (HMD) are used to visualize AR content. We are observing a trend in using HMDs especially for procedural tasks because they allow working hands-free. Due to several limitations regarding business, technology and human aspects, this technology is not being used daily. In order to successfully implement such a new technology into the industrial field and other domains, all three aspects ought to be addressed. Characteristic key performance indicators (KPIs) provide measures to demonstrate clear advantages, such as efficiency and quality improvements, in contrast to established methods. In addition to that, a human-centered design approach is necessary to address relevant human demands by executing interviews and field observations [2]. The technology must be evaluated under real conditions in order to conceive all influencing factors, however most evaluations are conducted in limited laboratory environments with low complexity LEGO Duplo assembly tasks [3]. In this paper we aim to overcome these limitations by making the following contributions:

- In chapter two we give an overview about traditional and AR-based assembly training. We also present some limitations which we want to outline in this paper.
- We developed an application for an engine assembly training task using the Microsoft HoloLens HMD. Necessary hardware and software details are presented in the third chapter.
- We execute a user study with 36 participants using an engine assembly training task. Comprehensive details about the study design, the setup, the procedure and the participants are given in chapter four.
- The results are shown in chapter five. During the study we measured the time as well as picking and assembly mistakes in a subsequent intermediate recall. We also assess the software usability using the system usability scale (SUS).
- The sixth and last chapter summarizes this paper and describes our future steps.

2 State-of-the-Art

In this chapter we provide an overview about traditional assembly training in the automotive industry and new learning approaches using AR-based technologies.

2.1 Assembly Training

The apprenticeship of new employees is one of the crucial aspects to ensure high quality processes and products. Assembly and maintenance training aims to acquire procedural as well as fine-motor skills [4]. Trainees have to learn how to assemble a specific object in a predefined order and cycle time, using the correct parts and tools.

Depending on the scope of work, this skill-acquisition usually takes several days. The underlying job instructions are based on a four-step methodology (Fig. 1) invented by the training-within-industries (TWI) institute, which was founded during the Second World War [5]. The main purpose at that time was to train millions of unemployed people to increase the production outcome of military goods such as weapons, tanks and warplanes. The concept of TWI was adapted by the Japanese government at the end of 1945 to reconstruct their industries. It became an integral part of the Toyota Production System and a role model for the entire automotive industry up until today.

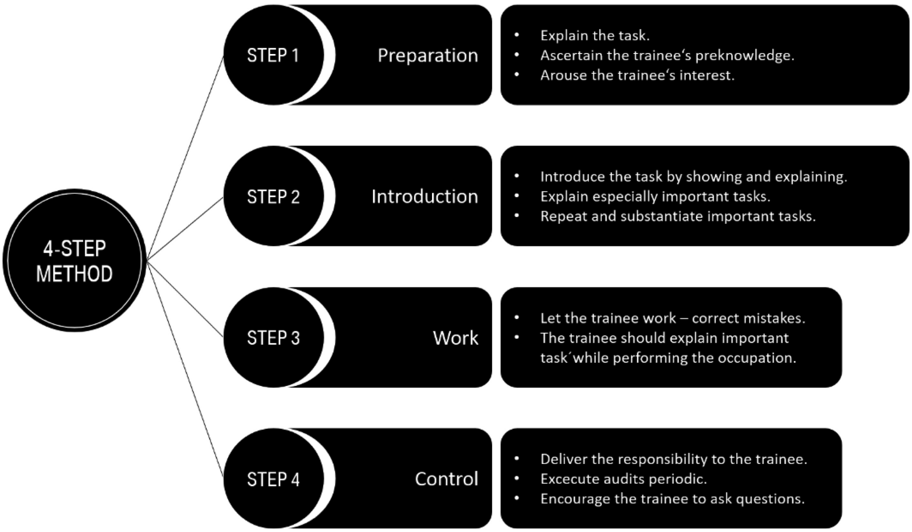


Fig. 1. The 4-step methodology for work instructions.

The first step aims to explain the task in detail while creating a pleasant work environment to make him feel comfortable. The trainer further arouses the trainee's interest and ascertains any prior knowledge. The second step represents the introduction to the task itself. Demonstrating the task whilst showing and explaining important details, the trainer provides a first overview and understanding of the occupation to the trainee. In the third phase, the trainee starts working on and explaining the tasks while the trainer pays attention and makes corrections in case of any mistakes. After repeating the work process several times, the trainee is able to work independently. The trainer delivers full responsibility to the trainee and executes audits at regular intervals.

This methodology is a standard procedure in the automotive industry and other industrial domains but not contemporary. Due to increasing globalization and several causes for migration, the world population is mixing. Naturally, this social phenomenon also affects the entire workforce of a company. Employees of different origin and various language backgrounds are required to work together in the same work environment. Depending on the extent of a possible language barrier and the trainer's pedagogical abilities, the learning time and success vary between employees. In addition to that, a trainer is assigned to one and the same trainee until the training is completed, which does not add value and should be reduced to a minimum in order to save human resources. Furthermore, assembly training is realized during the actual production at the assembly line, which provokes stress because of short cycle times. Most trainees are not able to assemble a specific object in the given cycle time at the beginning of a training session. The learning process takes mostly three to four days, depending on the task complexity [2]. Due to these reasons, assembly training needs to be optimized. Therefore, AR is being introduced to the industrial domain and promises to increase the efficiency and quality of procedural tasks. We present respective approaches in the following Sect. 2.2.

2.2 AR-Based Training

Holm mentioned that it "(…) *would be really good if we could create a working environment looking more like a video game"* [6]. Especially for young people, who will be the future shop-floor operators, such fascinating work environments become more and more interesting. Established technologies such as laptops, smartphones and projectors as well as new technologies such as head-mounted-displays (HMD) can support this idea by visualizing augmented data into our real world. These approaches not only affect our daily life, they also promise to increase the efficiency and quality of procedural tasks, such as assembly training tasks. Previous studies reinforcing this potential are presented in the following paragraphs.

A mobile solution using a tablet, was developed by Rios [7] to enhance the maintenance procedures of an Engine Air Bleed System on the Boeing 737. A study was executed with 34 students comparing the impact of written instructions to those of AR features. The assembly time and assembly quality using a 46 point scale were measured. Participants performed 17.22% slower when using the AR learning approach but reached a quality increment of 24% compared to the traditional method.

Another study [8] aimed to compare the effects of paper-based instructions to a screen-based solution, in which augmented instructions were presented on a screen at the workplace in front of the participants. Twenty volunteers participated in each experimental group while the assembly time was measured. Results indicated a two minute time improvement for the first attempt of assembling a gully trap compared to the paper-based method. Participants in the AR group further were able to fully assemble the gully trap after ten attempts; the other group needed twelve attempts, which indicates quicker learning success when using AR.

An experiment with twenty participants compared HMD-based training to face-to-face training [9]. A modified Oculus Rift HMD equipped with two additional cameras was used as a video-see-through HMD. An aircraft maintenance task was chosen for this study. Twenty-four volunteers were randomly assigned to one of the two conditions using a between subject design. Trainees had to perform a multiple choice questionnaire in order to assess how much knowledge was retained during the training session. An additional second evaluation aimed to gather the captured knowledge. All of the participants were asked to perform the task step by step until completing the entire operation. Results indicated no significant differences between both tests.

More evaluations using AR for assembly and maintenance training tasks are presented in comprehensive contemporary surveys [10]. Although its potential has been proven promising and the advantages seem obvious, AR is not in practical use in the automotive industry due to several scientific limitations [11].

- Most of the assessments examining AR are realized under laboratory conditions.
- Participants are mostly students without prior knowledge of assembly tasks.
- Procedural tasks do not match the complexity of industrial assembly tasks.
- Reliable comparisons between HMD-based and face-to-face training are still missing.
- Researchers are rather measuring the training time instead of gathering the training transfer, which is the main KPI for training assessment.
- There are only a few scientific contributions evaluating the advantages of HMD-based assembly training in industrial settings.

These missing contributions, along with hardware and software challenges, make it difficult to transfer, adapt and scale the established knowledge to real industrial use cases. A profit-oriented domain such as the entire automotive industry is not inclined to use a technology without a clearly proven benefit. Hence, in this paper, we provide a first approach to overcome this deadlock.

2.3 Lean Approach

To overcome the uncertainty of how to present information within AR based trainings, the whole teaching process as well as the visual and auditory information that is being presented in the Microsoft HoloLens in this study is based on lean principles (Fig. 2). The use of lean principles enables manufacturing processes as well as whole organizations to achieve increasingly high levels of efficiency and to compete at a low cost

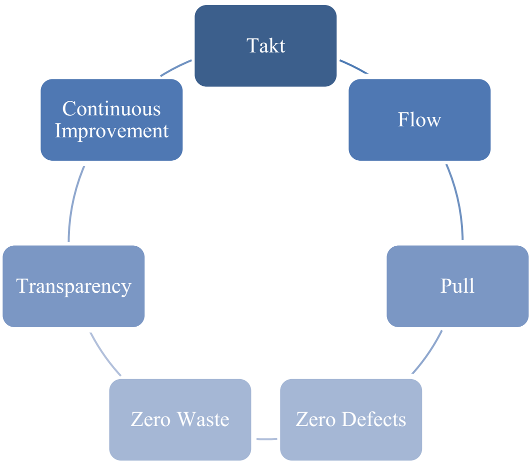


Fig. 2. Lean principles

level, while maintaining high speed of delivery and gaining optimum quality [12]. Eight main principles are taken into consideration for the creation of the AR application used in this study (Table 1).

Table 1. Lean principles applied for the AR application

Takt	The training with the application can be performed at an individual speed level. The trainee controls the time he or she needs for the training steps autonomously (takt time)
Pull	The trainee is able to get the necessary information at the time he or she needs it. No information is pushed in the process as the trainee retrieves the subsequent information about the task autonomously by saying "next" or "back". Thus, the information is pulled by the demand of the trainee
Flow	By using a step-by-step presentation of the task with visually and auditory congruent information, the trainee is given a one-piece flow of information
Zero defects	Only the correct order, position and process are being demonstrated. Showing how to do it incorrectly can increase the failure rate
Zero waste	Constant transparency of all process steps with all necessary information reduces the time for questions and errors. Every step can be repeated as many times as desired. This creates an individual, autonomous training session
Transparency	Transparency of all process steps and all teaching content is provided. No information is left out. All information can be repeated autonomously
Standards	Every trainee receives the same information in the same order along with the same visual and auditory stimuli. The best practice is provided to all trainees
Continuous improvement	Trainees can improve their performance immediately while completing the task. No model learning is necessary because there is immediate learning by trial and error

If AR applications are construed taking these eight principles into consideration, procedural tasks will more likely match the complexity of industrial assembly tasks.

3 Technology

In this chapter we describe the hardware and software used in our experiment.

3.1 Hardware

We used a Microsoft HoloLens (HL) which is an optical see through HMD with a field-of-view (FOV) around 30 * 17°. The HL allows enriching the real environment with 3D virtual information. It works standalone with a 32 bit architecture without any additional computer and includes a CPU, GPU and HPU. Several sensors constantly collect data about the internal and external environment to synchronize the AR world with the real world (Fig. 2).

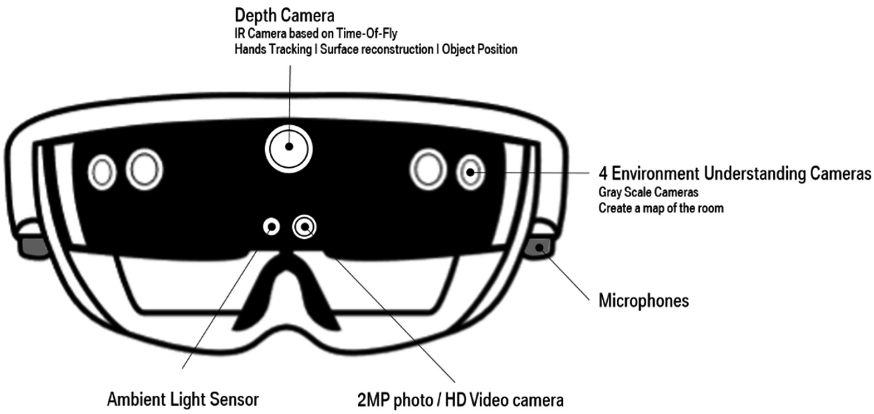


Fig. 3. Microsoft HoloLens front view sensor overview.

The inertial measuring unit (IMU) composed of an accelerometer, a gyroscope and a magnetometer measures the velocity, orientation and gravitational forces of the HL. Therefore, it is responsible for the three degrees of freedom (DOF) and the six DOF tracking system. The tracking systems consist of 3 DOF rotation and 3 DOF translation tracking to ensure a truly immersive AR experience. A depth camera in front of the HL is used to capture the gesture tracking (input data). Four more cameras, two on each side, are used to gather the room orientation what is known as inside-out tracking. The ambient light sensor is measuring the light intensity in the surroundings and adjusts the hologram brightness accordingly. An additional 2MP photo/HD video camera is integrated in the front of the HMD. The four microphones are used to capture voice input and display audio files.

3.2 Software

We developed a software for the HL using the game engine Unity3D. This game engine offers several advantages for HL application development. Microsoft provides a software development kit (SDK) which is fully compatible with Unity3D and therefore, it is easy to create immersive AR applications. We also used Photoshop for our user interface (UI) design. Our application followed a four level structure which is shown in Fig. 3.

At the beginning of the application, the user gets a short audio introduction, combined with a welcome hologram visible in his FOV (1.1). The user can choose between three different work cycles (1.2) by either air tapping (gesture interaction) or by voice command. Every work cycle has a different work content. Once the user selects one of the three work cycles, he is asked to walk to the respective workstation.

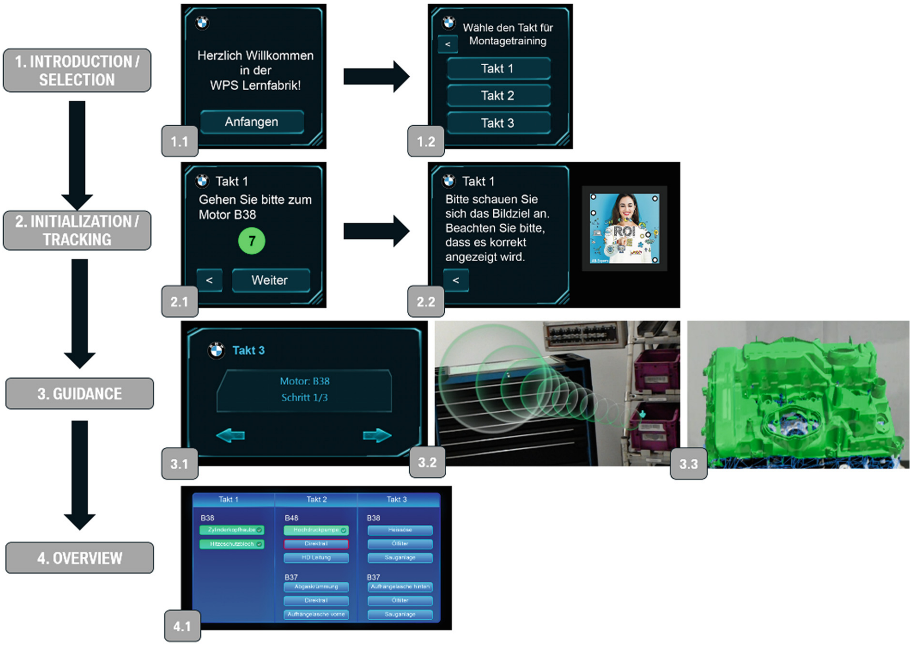


Fig. 4. Software level structure.

As soon as the user arrives at the given destination, he can go to the next step by either air tapping "weiter" or by using the voice command next (2.1). The initialization follows immediately afterwards. The user is encouraged to track the given imagine displayed in the hologram by looking at it (2.2). We used Vuforia image-based tracking for the initialization and placed the image target relative to the engine by using a photogrammetry approach to realize a sufficient overlay. The superimposition stays unaltered as long as the distance between the image target and engine remains unchanged. The training starts after correct initialization. We implemented an instruction panel (3.1), which is always visible above the engine, where the user receives text-based instructions about the current task. The user can go further or back by either air tapping the arrows or by using the voice commands next and back. We further implemented a Bezier curve (3.2) which was recommended in [13] as a best practice approach for guidance at picking tasks. The user receives text information in the instruction panel, additional audio guidance and the Bezier curve to pick the correct part from the shelf. This multimodal learning approach was recommend in previous studies [14]. By saying "next", the user gets to the next visual feature which is a 3D virtual overlay (3.3) and helps to understand the part orientation and location. After successfully assembling the part at the predefined location, the user can continue with the next step (Fig. 4).

Building a mental representation of a given procedural tasks was found to be very important for the learning success [4]. Therefore, we implemented an overview panel

(4.1) where the user receives information about finished, current and upcoming tasks. This holistic overview can help strengthening the mental model building.

The entire software was developed based on previous comprehensive demand analysis [2], general software design guidelines according to the DIN EN ISO 9241-110 [15] and domain specific recommendations [16, 17]. We evaluated this software product in order to find weaknesses and compare this learning approach to the established face-to-face training. The next chapter describes our methodology in detail.

4 User Study

We conducted a user study with 36 participants to compare the effects of HMD-based assembly training to the traditional face-to-face training. This section describes the study design, explains the procedure, introduces the hardware setup, gives detailed information about the participants and reports the results of our measurements.

4.1 Design

We used a repeated measure experiment as a study design with one independent variable, i.e. instructions given by a person (group A) compared to instructions given by an augmented reality application (group B) to test the following hypotheses:

H0: There is no significant difference between face-to-face training and AR–to-face training.

H1: There is a significant difference between face–to-face training and AR-to-face training

As dependent variables, we considered instruction time, quality (measured by correctly installed parts, correct order of installed parts, picking mistakes, maintaining the required cycle time (45 s) and forgotten parts), system usability scale (SUS) according to Bangor [18] and gained knowledge about the task (questionnaire).

4.2 Apparatus and Setup

The workplace used for the setup is the VPS learning factory with an assembly line of seven engines. We chose an identical task description for each instruction system (face-to-face, AR-to-face) to ensure that the instructions as well as the tasks feature the same complexity level. The trainees assembled the engines B37, B38 and B48. The workplace is formed by three assembly spots and automated guided vehicles (AGV) which carry the engines on top of them.

The following figure displays the training situation viewed from above. There are three assembly training spots (AP1, AP2 and AP3) and two disassembly spots (DM1, DM2). The engines rotate on the AGVs. For visualization, imagine a B38 is placed in front of AP1 where the first trainee assembles all the parts in 45 s, afterwards this engine moves in front of AP2, where another trainee assembles the B38, and finally the

AGV moves to AP3 and the trainee on this spot assembles the final parts. Finally, the engine B38 is completely assembled. Analogous to this logic, B48 and B37 are assembled (Fig. 5).

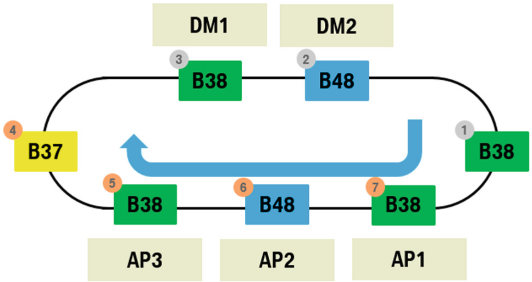


Fig. 5. Working space

This user study is divided into two parts. In the first part, trainees received instructions on how to assemble the engines, whilst in the second part, trainees were encouraged to assemble the engines without any assistance. In one group (A), three trainees were provided with instructions on how to assemble the engines by three trainers at the same time, while allocated to their assembly spots. In another group (B), trainers were replaced by using the application in the HoloLens. Afterwards, the trainees assemble the engines at the same time following the logic described above.

4.3 Procedure

In consequence of the amount of participants, the study was conducted on two different days. The procedure was the identical on both days. After informing the participants about the main interest of the study and obtaining consent, demographic information (age and sex) was collected. Group A and group B each had a maximum of twenty minutes to learn how to assemble the engines during the instruction part. Group A was given the relevant task information by a trainer face-to-face. Group B received all relevant task information by an augmented reality application using the Microsoft HoloLens as mentioned above. While the instructions for group A and group B were identical, group B was given an additional 10 min to learn how to handle the device before the training session started. After the training session, each of the two groups assembled the engines without any assistance for fifteen minutes. During the assembly, the number of correctly installed parts, the correct order of installed parts, picking mistakes, maintenance of the required cycle time (45 s.) and number of forgotten parts were measured.

After the training and assembly session, participants completed the SUS and a questionnaire retrieving the gained knowledge about the task. Fisher´s exact test was used to analyze the data (Figs. 6 and 7).

Fig. 6. Instructions by trainer (group A)

Fig. 7. Instructions by HoloLens (group B)

4.4 Participants

We recruited 36 participants (35 men, 1 women), aged from 16 to 26 (M = 19.91 SD = 2.31). All participants are production mechanic trainees working for an automotive company with no prior experience in using or working with AR. Informed consent was obtained from all participants.

5 Results

During the training we measured the time needed for installing the required parts. The results showed that assembly training takes more time when using a HMD (H1). Trainees needed 60% more time to complete the training task (Group A: M = 9 min 40 s; Group B: 15 min 51 s) in the HMD condition compared to the trainees in the face-to-face condition. This result was to be expected, because trainees first needed to familiarize with the gesture and voice interaction methodology. Once users become used to the HMD, their performance speed is expected to increase. During the immediate recall, we measured numbers of picking mistakes and assembly order mistakes. Results showed that trainees who performed the assembly task with HMD support made 10% less picking mistakes, 5% less assembly order mistakes and caused 60% less rework.

During the assembly session, we measured the quality of the training outcome by considering the number of correctly assembled parts. The data analysis showed that there is a significant difference between being trained by a trainer and being trained by AR. The overall quality (Qg) and the quality per tact (Qt1, Qt2, Qt3) is higher when a trainer has giving the information. The reason for this outcome is the fact that essential parts of the different tacts were not assembled correctly in the AR condition (HSB, HDP, HDL, HÖ) and therefore overall quality and the quality per tact were higher when the information was given by a trainer (Table 2).

Splitting the dataset according to the two study days (day1/day2), it becomes apparent that on day one, there was no significant difference between being trained by a trainer or being trained by AR, whereas on day two, there is a significant difference between being trained by a trainer and being trained by AR (Tables 3 and 4).

Table 2. Significance test

	N	Fisher´s exact test (p)	Significant difference?	Which method is more effective? (Trainer/AR)
Qg * Group	132	0.009	Yes	Trainer
Qt1 * Group	132	0.011	Yes	Trainer
Qt2 * Group	144	0.004	Yes	Trainer
Qt3 * Group	156	0	Yes	Trainer
ZKH * Group	132	1	No	Equal
HSB * Group	108	0.007	Yes	Trainer
HDP * Group	120	0.097	Yes	Trainer
DR * Group	144	0.49	No	Equal
HDL * Group	120	0.037	Yes	Trainer
AK * Group	24	0.4	No	Equal
AL * Group	24	1	No	Equal
HÖ * Group	156	0.00	Yes	Trainer
ÖF * Group	156	0.183	No	Equal
SM * Group	156	0.198	No	Equal

Table 3. Significance test day 1

	N	Fisher´s exact test (p)	Significant difference?	Which method is more effective? (Trainer/AR)
Qg * Group	66	0.427	No	Equal
Qt1 * Group	66	0.606	No	Equal
Qt2 * Group	72	0.245	No	Equal
Qt3 * Group	78	1	No	Equal
ZKH * Group	66	1	No	Equal
HSB * Group	54	0.583	No	Equal
HDP * Group	60	1	No	Equal
DR * Group	72	0.144	No	Equal
HDL * Group	60	0.143	No	Equal
AK * Group	12	1	No	Equal
AL * Group	12	1	No	Equal
HÖ * Group	78	1	No	Equal
ÖF * Group	78	1	No	Equal
SM * Group	78	1	No	Equal

After finishing the immediate recall assembly-cycle, trainees were further asked to accomplish a knowledge retention test consisting of four questions. This test aimed to capture the knowledge acquisition of the performed assembly task. The knowledge retention test indicated that trainees using HMD gathered the same knowledge as participants trained face to face (Group A: 94% and Group B: 93% correct answers). The HMD training group was asked to additionally fill out the system usability scale

Table 4. Significance test day 2

	N	Fisher´s exact test (p)	Significant difference?	Which method is more effective? (Trainer/AR)
Qg * Group	66	0.011	Yes	Trainer
Qt1 * Group	66	0.005	Yes	Trainer
Qt2 * Group	72	0.009	Yes	Trainer
Qt3 * Group	78	0	Yes	Trainer
ZKH * Group	66	1	No	Equal
HSB * Group	54	0.002	Yes	Trainer
HDP * Group	60	0.015	Yes	Trainer
DR * Group	72	0.806	No	Equal
HDL * Group	60	0.195	No	Equal
AK * Group	12	0.545	No	Equal
AL * Group	12	1	No	Equal
HÖ * Group	78	1	No	Equal
ÖF * Group	78	0	Yes	Trainer
SM * Group	78	0.013	Yes	Trainer

(SUS) questionnaire. This questionnaire featuring ten items aims to rate the usability of our training software to examine whether the HMD-based application fulfills required acceptance standards for software by the user. Participants were asked to score the items with response possibilities ranging from "strongly agree" to "strongly disagree". Scores from zero to one-hundred allowed rating the usability of our system, whereby 100 meant 'best imaginable'. Results indicated an average SUS of 73,5 which means 'good'. Overall, our HMD-based application constitutes a suitable option for assembly training but still has potential for improvement.

6 Conclusion

In this study, we compared two different methods of how to train an assembly task (face-to-face vs. AR-to-face). We build an application for a real engine assembly training task and tested the effects on time, quality, picking mistakes, assembly order mistakes, rework and knowledge retention. The significant quality differences between the two study days could be traced backed to several influencing factors. First, possible process instabilities with regard to the training process itself could have caused these differences. Using HMDs influences the training preparation, training conduction and training follow up. Furthermore the face-to-face training process is well known by the trainer, since it is performed for several years, where else the AR training process is only performed since a few months. Therefore, not every process step in the AR training is performed as fluently as in the face-to-face training. Second, individual differences in using communication and information technology and the personal motivation in using new technologies could also have an effect on the variable quality.

In conclusion, it seems AR-based training can be an alternative to face-to-face training but still remains some weaknesses that could not entirely be overcome by including lean principles into to the application. Considering future work, we want to improve our application by providing more user control, by adding gamification elements, by offering different user modes ranging from a tutorial level to an expert level, by using multimodal information as well as various feedback modalities. Furthermore, we plan to conduct more user studies using a more complex assembly training tasks.

References

1. Syberfeldt, A., Holm, M., Danielsson, O., Wang, L., Brewster, R.L.: Support systems on the industrial shop-floors of the future – operators' perspective on augmented reality. In: Procedia CIRP, pp. 108–113 (2016). https://doi.org/10.1016/j.procir.2016.02.017
2. Werrlich, S., Nitsche, K., Notni, G.: Demand analysis for an augmented reality based assembly training. In: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 416–422 (2017). https://doi.org/10.1145/3056540.3076190
3. Loch, F., Quint, F., Brishtel, I.: Comparing video and augmented reality assistance in manual assembly. In: 2016 12th International Conference 2016, pp. 147–150 (2016). https://doi.org/10.1109/IE.2016.31
4. Webel, S., Bockholt, U., Keil, J.: Design criteria for AR-based training of maintenance and assembly tasks. In: Shumaker, R. (ed.) VMR 2011. LNCS, vol. 6773, pp. 123–132. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22021-0_15
5. Graupp, P., Wrona, R.J.: The TWI workbook: essential skills of supervisors (2016). ISBN 9781498703963
6. Holm, M., Adamson, G., Moore, P., Wang, L.: Why I want to be a future Swedish shop-floor operator. Procedia CIRP **41**, 1101–1106 (2016). https://doi.org/10.1016/j.procir.2015.12.057
7. Rios, H., González, E., Rodriguez, C., Siller, H.R., Contero, M.: A mobile solution to enhance training and execution of troubleshooting techniques of the engine air bleed system on boeing 737. Procedia Comput. Sci. **25**, 161–170 (2013). https://doi.org/10.1016/j.procs.2013.11.020
8. Hořejší, P.: Augmented reality system for virtual training of parts assembly. Procedia Eng. **100**, 699–706 (2015). https://doi.org/10.1016/j.proeng.2015.01.422
9. Gonzalez-Franco, M., Pizarro, R., Cermeron, J., Li, K., Thorn, J., Hutabarat, W., Tiwari, A., Bermell-Garcia, P.: Immersive mixed reality for manufacturing training. Front. Robot. AI **4**, 3 (2017). https://doi.org/10.3389/frobt.2017.00003
10. Wang, X., Ong, S.K., Nee, A.Y.C.: A comprehensive survey of augmented reality assembly research. Adv. Manufact. **4**, 1–22 (2016). https://doi.org/10.1007/s40436-015-0131-4
11. Werrlich, S., Eichstetter, E., Nitsche, K., Notni, G.: An overview of evaluations using augmented reality for assembly training tasks. Int. J. Comput. Inf. Eng. **11**(10), 1068–1074 (2017)
12. Bhamu, J., Singh Sangwan, K.: Lean manufacturing: literature review and research issues. Int. J. Oper. Prod. Manag. **34**, 876–940 (2014). https://doi.org/10.1108/IJOPM-08-2012-0315
13. Schwerdtfeger, B.: Pick-by-vision: bringing HMD-based augmented reality into the warehouse (2012). ISBN 978-3832526276

14. Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., Preusche, C.: An augmented reality training platform for assembly and maintenance skills. Robot. Auton. Syst. **61**, 398–403 (2013). https://doi.org/10.1016/j.robot.2012.09.013
15. Ergonomics of human-system interaction - Part 110: dialogue principles (ISO 9241-110:2006). Deutsche Fassung EN ISO 9241-110:2006
16. Regenbrecht, H., Baratoff, G., Wilke, W.: Augmented reality projects in the automotive and aerospace industries. IEEE Comput. Graph. Appl. **25**, 48–56 (2005). https://doi.org/10.1109/mcg.2005.124
17. Funk, M., Kosch, T., Kettner, R., Korn, O., Schmidt, A.: motionEAP: an overview of 4 years of combining industrial assembly with augmented reality for industry 4.0 (2016)
18. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Hum.-Comput. Inter. **24**, 574–594 (2008). https://doi.org/10.1080/10447310802205776

Author Index