VWSocialLab: Prototype Virtual World (VW) Toolkit for Social and Behavioral Science Experimental Set-Up and Control

Lana Jaff[1], Austen Hayes[2], and Amy Ulinski Banic[1]

[1] University of Wyoming, Dept. 3315, 1000 E. University Ave, Laramie WY, 82071
[2] Clemson University, 100 McAdams Hall, Clemson, S.C. 29634
{ljaff, abanic}@uwyo.edu, ahayes@clemson.edu

Abstract. There are benefits for social and behavioral researchers to conduct studies in online virtual worlds. However, typically learning scripting takes additional time or money to hire a consultant. We propose a prototype Virtual World Toolkit for to help researchers design, set up, and run experiments in Virtual Worlds, with little coding or scripting experience needed. We explored three types of prototype designs, focused on a traditional interface with pilot results. We also present results of initial expert user study of our toolkit to determine the learnability, usability, and feasibility of our toolkit to conduct experiments. Results suggest that our toolkit requires little training and sufficient capabilities for a basic experiment. The toolkit received a great feedback from a number of expert users who thought that it is a promising first version that lays the foundation to more future improvements. This toolkit prototype contributes to enhancing researchers' capabilities in conducting social/behavioral studies in virtual worlds and hopefully will empower social and behavioral researchers by proving a toolkit prototype that requires less time, efforts and costs to setup stimulus responses types of human subject studies in virtual worlds.

Keywords: Virtual Humans, Online Virtual Worlds, Virtual Environments, Social Science, Psychology, Behavioral Science, Human Experiments, Toolkit, Evaluation, Prototype Experimental Testbed.

1 Introduction and Motivation

Online Virtual Worlds, such as second life and 3rd Rock Grid (3RG), have been widely used for educational and entertainment purposes [2,7,9,12]. Recently, these environments have also been sufficiently used as platforms to conduct studies and experiments in various fields [3,5,10]. Virtual Worlds have economic and political systems that provide interesting social dynamics that have been studied by researchers in social and behavioral sciences, yet not fully explored. Due to the computational nature of these virtual worlds, they lends themselves to be used as virtual laboratories for conducting human subjects experiments in the social and behavioral sciences, typically found in fields such as Sociology and Psychology, since many of the

R. Shumaker (Ed.): VAMR/HCII 2013, Part I, LNCS 8021, pp. 39–48, 2013.

environmental and social variables can be controlled, and automated data collection tools can be included. Over the years, social and behavioral researchers have shown great interest in using Virtual Worlds as platforms for conducting their studies [3,5,8,10]. Psychologists have expressed that virtual worlds increase participants' "engagement" and their reactions, therefore may increase the reliability and effects of the experiments on the participants [3]. Furthermore, virtual worlds may minimize the "lack of replication" and "sampling problems" found in traditional laboratories [2,7,10,13]. However, in previous studies conducted in Virtual Worlds, scientists had to either learn the scripting skills to set up an experiment or collaborate with other professionals, which may be time consuming and costly. The objective of this research is to design a method to empower researchers to be able to set up and conduct experiments in Virtual Worlds without needing to learn scripting and depend on other professionals. The goal of this research was to use virtual worlds as an alternative platform to design and conduct standard social and behavioral science experiments.

We designed and developed all the components needed to conduct an experiment in a Virtual World that involved a simple task, mixed design of between and within subject conditions, a controlled virtual character involved in the experiment, and multiple data types collected. Additionally we interviewed several social and behavioral science researchers to learn about their experimental goals and limitations. As a result, we propose a prototype Virtual World Toolkit for researchers to use to help design, set up, and run experiments in Virtual Worlds, with little coding or scripting experience needed. We explored three types of prototype designs: one visual with drag and drop features, one conversational where a virtual human discusses the experimental study and then implements it, and the third a more traditional button-like interface. Our hypothesis was that our toolkit is easy to use, has sufficient capabilities to set up and conduct a basic human subject experiment, and requires little training and less time, as opposed to learning scripting skills, to get started. We also hypothesized that our toolkit will receive positive attitude ratings similar or better than the other toolkits that the researchers currently use to conduct their studies.

2 Background and Related Work

There have been a number of studies conducted in Second Life virtual world to evaluate the effects of the presence of virtual characters on task performance. In real world social and behavioral studies, there is a theory that refers to the effects of presence of others (real humans) during task performance called social facilitation/inhibition [4,8,14]. This theory states that the presence of others affects the performance of novel tasks more than the performance of learned tasks [4,8,14]. Participants perform better on learned tasks and worse on novel tasks. A number of studies were conducted in virtual environments (including online virtual worlds such as Second Life) and aimed to study the social facilitation/inhibition with the use of virtual humans as the audience while real participants perform different levels of tasks [4,8,13,14]. Hayes conducted a study to evaluate the social facilitation/inhibition effects among simple and complex tasks [8]. This study found that the social facilitation/inhibition theory applies to virtual worlds where the results showed that the participants were affected by the presence of the virtual observers (male or female) during the performance of the complex tasks. Yee conducted another study to evaluate the effects of the avatar's gender, distance and eye-gaze on the social behaviors between avatars in virtual environments [11]. The results of this study showed that male avatars tend to keep a larger personal distance and less eye contact with other male avatars, while female avatars keep a smaller personal distance with other female avatars and more eye contact. Antonio investigated the connection between avatar's behaviors with one another in virtual worlds like eye contact, conversations and the application of "clustering techniques" [1]. The results were applied to real life relationships between teachers and students by better understanding the social behaviors of students in class. The data collected from the study showed that some students were paying attention, especially with eye contact, to lecture while others did not.

3 Preliminary Research: Exploring Components Needed to Conduct Social and Behavioral Studies in Virtual Worlds

For our preliminary research, we determined the necessary components to conduct a simple stimulus-response human-subjects studies. We designed the toolkit to allow for flexible options for experiments to identify independent variables, such as the between and within subject conditions, and dependent variables, such as the automated data collected, trials, orders and animations, for automated virtual character control. The toolkit takes advantage of the scripting capabilities of the virtual world, yet abstracts that from the researchers by presenting buttons and menus to interact with rather than scripting the components. We developed the toolkit on a 3D Rock Grid leased island (3RG). Our prototype was based on a social-facilitation study with virtual human audience types. Researchers can use and interact with the toolkit using an avatar, or a computer generated character controlled by a human. The capabilities are:

- Number of conditions: Consists of an empty edit field where the number of conditions is entered using the keyboard.
- Manipulation of conditions: Select one of two options to manipulate the study conditions (within subjects, between subjects) by clicking on an option.
- Observer Avatar: Consist of three options as they relate to between subjects conditions. This data will be sent to the data collection note card for reference.
- Participant Avatar: Consist of two options: Male and Female. This data will be sent to the data collection note card for reference as they relate to the experimental participant that will take part in the study when researchers set up their experiments.
- Number of trials: Number of trials is entered using the keyboard.
- Order of trials: Select one of three options that represent the order in which the trials are presented in (randomize, Specific order, Balanced Latin squares) by clicking on an option using the left mouse button.

- Response time: Includes three options to select how to record the response time (per block, per trial and per condition) by clicking on an option.
- Completion time: Includes three options to select how to record the completion time (per block, per trial and per condition) by clicking on an option.
- Task accuracy: Defines how to measure accuracy of the tasks using the keyboard.
- Task errors: Defines what the errors of the tasks are using the keyboard.
- Input Method: What is used as an input method to respond to the tasks? Keyboard is the only input method provided for the toolkit.
- Task output: What is used as the output stimuli (to provide feedback for the user's responses)? Two options are provided: Textures and Play sound.
- Start: This is the last button to press in order to start the study after setting up the above components by clicking on this button using the left mouse button.
- Avatar Appearance: Provides four common and required appearances to conduct studies in virtual worlds: professional, casual, hot-trendy looks and bold (rock and roll) looks. For each of these appearances, we provided different body shapes (tall, short, skinny, muscular), skin colors (white,dark,tan,male base), eye colors (green,blue,brown,grey) and hair styles including bald.
- Avatar Gestures: Provides 10 common gestures for the observer avatar that observes the participant's avatar during task performance. These gestures include sad, angry, impatient, embarrassed, laugh, unhappy, wave, worried, cough and bored.
- A task display board: This board displays the tasks of the set up study for the participant after pressing the "Start" button by the researcher.

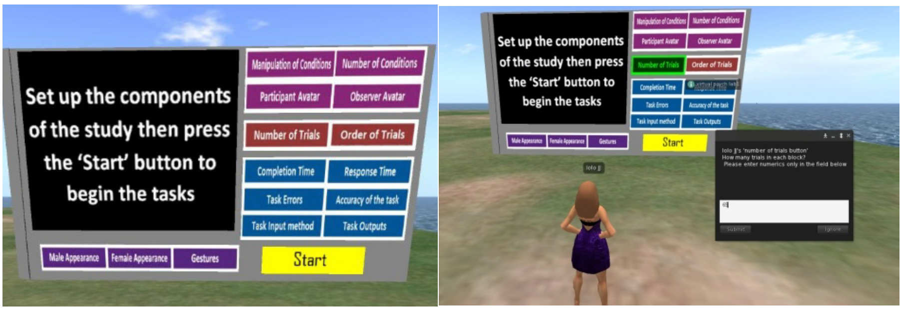

Fig. 1. Example of the prototype interface for parts of the toolkit in the Virtual World

4 Experimental Design

4.1 Physical and Virtual Environments

All participants completed experimental tasks and questionnaire in a physical testing room at L41 lab of the Engineering building at the University of Wyoming. We used an Intel Core 2, 2.66 GHz, 2 GHz RAM Dell PC with an ATI Radeon HD 160 graphics card attached to a 20 inch at screen monitor to display the virtual world. In addition, we used a different PC (Intel Core (TM) 2.80 GHz, 2 GHz RAM Dell PC with an NVIDIA Quadro FX 580 graphic card attached to a 20 inch at screen) for completing the questionnaires. Zen Viewer version 3.4.1 was used to view the 3RG virtual world. The participants were given a gender-appropriate avatar to control and interact with our toolkit. Our toolkit was created in a 3RG leased island. It consists of a collection of objects integrated together and scripted using Linden scripting language (LSL). A gender appropriate avatar (called the observer avatar) was given to the participant to use to setup studies using the toolkit. Another avatar was given to the participants (called the experimental avatar) and the gender was varied for this avatar. The experimental avatar was used to play the role of an experimental participant and perform that tasks of the studies set up in tasks A and C.

4.2 Experimental Tasks

Our experiment consisted of three tasks that were manipulated within subjects.

Task A. We asked the participants to set up the components of a specific stimulus-response type of study. We were assessing usability in this task. The participants were given specific math task components and were asked to set up the study using our toolkit without training learnability. For this task, we recorded the order in which the participants set up the study components as well as the number of presses used. After completing the tasks of this study, we showed the participant how to access and view the data collected from the study.

Task B. Participants created gestures and adjusted appearance for an avatar. The first level of task B included using our toolkit without training to modify the avatar's appearances according to specific criterion (color of skin, color of eyes, body type and outfit) that were varied for each participant. This task also included creating two different gestures for that avatar. For the second level of task B, we gave the participant 10 minutes and asked them to figure out how to change the appearance of the avatar into a specific appearance (specific color of skin, color of eyes, body type and outfit) without using the toolkit. Afterwards we gave the participant another 10 minutes to create two specific gestures for that avatar without using the toolkit as well.

Task C. We asked the participants to set up a study of their own using the toolkit after giving them a short training on how to upload their own tasks into the toolkit. We were assessing our toolkit's capabilities and usability to set up stimulus/responses types of studies in this task. We had asked each participant to prepare and bring at most five files (due to time constrains) of basic stimulus responses type of study that they have conducted or familiar with in JPEG format to complete this task. We gave the participants a short training session on how to upload these tasks into the toolkit then asked them to set up the components of that study.

4.3 Experimental Measures and Procedure

Pre-experimental questionnaires collected demographic data (gender, age, ethnicity, computer and virtual worlds use level) and experimental background data. We asked participants to rate the usability and user experience criterion of the toolkits that they previously or currently use to conduct human subject studies and were assessed on a 7-point numerical scale (1=Not at all to 7= a great deal). The learnability questionnaire was given to the participants after completing task A. We also asked participants to complete a questionnaire asking them questions about their attitude and opinions towards other tookits (after task B, part 1) and towards our toolkit (after task B, part 2). We assessed the attitude responses on a 5-point Likert scale (1=strongly agree, 5=strongly disagree). Item responses on a final questionnaire (after task C) were used to determine whether our toolkit has sufficient capabilities to perform appearance modification and create gestures for the avatar in an easier, less time consuming method as appose to the traditional methods in these environments. This questionnaire collected responses on a 7-point numerical scale (1=Not at all to 7= a great deal). We also asked open-ended questions about the toolkit and recommendations. Participants also completed a post-experimental co-presence questionnaire, which refers to the extent the participants felt they were inside the virtual world and interacting with the avatars, on a 7 point numerical scale.

Prior to the experiment, we gained consent from each participant and asked them to complete the pre-experimental questionnaire. Participants were given brief training on how to use the avatar to move and interact with objects. After, we provided the list of components of the sample experiment and asked the participant to set it up using the toolkit. After completing task A, the participants completed the learnability questionnaire then proceeded to task B. The participants were asked to perform specific modifications on the observer avatar's appearance as well as creating gestures for it, then completed questions about the toolkit. For the second level of task B, the same observer avatar is used by the participants to modify its appearance, create a gesture for it without using the toolkit, and answer questions about the toolkit. The order of the task B parts was balanced. After completing all three tasks, the participants were instructed to move to complete the co-presence, usability and user experience questionnaires. Finally, the participants were debriefed and thanked for participation.

5 Evaluation Results

Mean (M) and standard deviation (SD) of learnability, usability and user experience percentages were computed by averaging across grouped questions for each participant in the pre and post experimental questionnaires. Order and number of presses of each button were computed by summing each item in the observer check sheet which is a list of observations while participation sheet that was used to record the quantitative data represented in the order and times each button in the toolkit was pressed. A paired samples (t-test) was conducted to test difference of means in comparing our toolkit against other applications used for conducting human subject studies and usability for virtual character control, where $p= 0.05$ was used to indicate significance. There were 6 expert participants, Faculty, PhD. and master students who conduct subject studies in the Psychology and computer science departments (however, not familiar with virtual worlds), males and females, and from the University of Wyoming. The mean age for the participants was 29.5, SD = 4.1 and they were randomly assigned a gender appropriate avatar.

5.1 Sufficient Capabilities

Our participants found that the toolkit meets the needs to set up standard stimulus responses types of studies, where (M = 5.3 , SD = 1.3) on a scale of 1-7. Results showed that the toolkit was rated as sufficient and high to set up stimulus responses types of studies, where (M = 6.5, SD = 0 .54). We asked the participants to write down the capabilities that our toolkit provides to set up and conduct stimulus response type of human subject studies in the post experimental questionnaire. These capabilities include in allowing as many conditions as necessary, allowing the setup of the study's tasks, trials and orders, as well as allowing multiple data collection tools and output stimuli. The results showed that the toolkit provides enough capabilities to setup stimulus response types of studies in virtual worlds where (M = 5.3, SD = 1.10). After calculating the mean and standard deviation for future usage and recommendation, the results showed that the participants are more likely to use the toolkit to set and conduct their future studies if they were to conduct studies in virtual environments and that they will recommend it to their peers. Where (M = 6, SD = 0.89) for likelihood of using the toolkit in the future studies in virtual environments and (M = 5.6, SD = 1.21) for recommending the toolkit to others.

5.2 Learnability, Usability and User Experience

The toolkit was rated positively by participants in regards to learnability, where M = 6.21 and SD = 0.49. Many users reported that they did not need programming/scripting experience to use the toolkit was calculated across participants, where M = 7, SD = 0. The results show that the participants did not need a lot of support to set up studies using the toolkit, where (M = 5.33, SD = 0.81). The results of a paired samples t-test to determine change in attitude from level 1 to level 2 of task B, showed that it is significantly easier to learn how to modify the avatar and create gestures using the toolkit without training, where $t(5) = 8.216$, $p < 0.001$,M = 4.60, and SD = 0.51 and $t(5) = 23$, $p < 0.001$,M = 4.80 and SD = 0.40 respectively, compared to learning how to modify the avatar and creating gestures using the traditional methods, without training, where (M = 1.6, SD = 0.81) and (M = 1, SD = 0.00) respectively.

As expected, the results show that setting up studies using the toolkit saves a significant amount of time, where (M = 6, SD = 0.89. The results show that the participants rated the toolkit as highly intuitive to set up stimulus responses types of studies where (M = 6.16, SD = 0.40). The participants rated the toolkit high in terms of consistency between other applications where (M = 6, SD = 0.63). The results of a paired samples t-test to determine change in attitude from level 1 to level 2 of tasks B, revealed that it is significantly easier to modify the avatar using the toolkit $t(5) = 19$, $p <$ 0:001,where (M = 5.0, SD = 0.0) compared to using the traditional virtual environments methods where (M = 1.8, SD = 0.16). It saves a significant amount of time to modify the avatar using the toolkit as expected, where $t(5) = 6.32$, $p = 0.001$,M = 4.50 and SD = 0.54 as opposed to using the traditional methods (M = 1.8,SD =0 .75). It is significantly easier to create gestures using the toolkit where $t(5) = 10$, $p < 0.001$,M = 4.60 and SD = 0.81 compared to the traditional methods, where (M = 1.3,SD = 0 .51). It was also found that creating gestures for the avatar using the toolkit saves a significant amount of time $t(5)= 5.94$, $p = 0.002$, $M = 4.6$ and $SD = 0.51$ if compared with the traditional methods of creating gestures in these environments, where ($M = 1.8$, $SD = 0.75$). The results showed that the participants rated the toolkit highly manageable to set up stimulus responses types of studies where ($M = 6$, $SD = 0.63$). The participants were satisfied with the toolkit, where $M = 5.8$ and $SD = 0.58$. The results showed that that participants were significantly satisfied with the toolkit, where $t(5) = 8$, $p < 0.001$,$M = 4.50$ and $SD = 0.83$ compared to their satisfaction ratings on the traditional methods in virtual worlds, where $M = 1.80$ and $SD = 0.40$.

5.3 Researchers Opinions: How Our Toolkit Compared to Other Similar Applications

A paired samples t-test results show no significant difference in comparing the toolkit's simplicity and ease of use to set up studies compared to other applications: $t(5) = -2.291$, $p = 0.071$, where ($M = 5.5$, $SD = 0.54$) for the toolkit and ($M = 3.3$, $SD = 2.16$) for other applications. After comparing the learnability of the toolkit compared to other applications, it was found that the toolkit is significantly easier to learn without training $t(5) = -2.557$, $p = 0.051$, where ($M = 6.3$, $SD = 0.51$) for the toolkit while ($M = 3.5$, $SD = 2.34$) for other applications. The results show that setting up studies using the toolkit saves a significant amount of time, $t(5) = -4.503$, $p = 0.006$, where ($M = 6$, $SD = 0.89$), compared to other applications ($M = 2.83$, $SD = 1.83$), and significantly more sufficient to set up studies, $t(5)= -4.108$, $p = 0.009$, ($M = 6.5$, $SD = 0.54$) as opposed to other applications ($M = 3.5$, $SD = 1.87$). The results show a significant difference in comparing the intuitiveness and manageability of the toolkit to set up studies compared to other applications where ($M = 6.16$, $SD = 0.40$ (for the toolkit), $t(5) = -3.955$, $p = 0.011$ and (for other applications) $M = 2.66$, $SD = 2.16$) and ($M = 6$, $SD = 0.63$ (for the toolkit), $t(5) = -3.162$, $p = 0.025$ and (for other applications)$M = 3.33$, $SD = 1.96$) respectively. Most of them made suggestions to enhance the toolkit and make it more suitable for their individual studies.

- "Brilliant idea for scientists in our field. I appreciate that I do not need to code”.
- "The design is similar to software I usually use and I don't think I need a manual".
- "Excellent idea which has the potential to make research easier in our field. I mostly liked the design which was pretty obvious and easy to use".

6 Discussion

The results show that our participants rated the toolkit to be practical to use. The results show that the toolkit provides enough capabilities to set up stimulus responses types of studies. Not all participants agreed to use the toolkit for all their current or future studies in general. However, all participants concurred to use the toolkit for current and future studies conducted in a virtual world, and that they would definitely recommend the toolkit to their peers. This fulfills the objective of our toolkit to be used as an alternative platform to setup experiments in virtual worlds. The participants were able to take advantage of the familiarity of the design with similar features to other applications that they usually use. The simple design of the toolkit has led the participants to complete the tasks for the first time without training. Many participants believed that they did not need scripting background or skills to use this toolkit to set up and conduct studies with accomplishes our objectives in creating a toolkit that is easy to learn and use with no or little scripting knowledge in the virtual world. The toolkit rated as significantly easier to use for avatars' appearance modification and gesture creation, than traditional methods. Our toolkit is limited to setting up stimulus responses types of tasks. Participants would like to add more capabilities, such as including adaptive types of task and response scales. In general, the toolkit meets the learnability, user experience and usability criterion. A number of expert users who thought that it is a promising first version that lays the foundation to more future improvements in order to make it more appropriate for setting up individual and more complex studies in virtual environments. The results show that the participants are more satisfied with our toolkit than others, though the researchers' opinions may be influenced by other factors that do not exist in our toolkit.

7 Conclusions, Contributions and Future Work

The results of this research have shown that the toolkit is easy to learn and use. It provides sufficient capabilities to setup stimulus response types of studies. The toolkit requires minimum training and coding skills and does not take too long to setup studies as opposed to learning the scripting skills. The results also showed that the toolkit provides alternative avatar control methods. The toolkit meets the learnability, user experience and usability criterion. The toolkit received a great feedback from a number of expert users who thought that it is a promising first version that lays the foundation to more future improvements in order to make it more appropriate for setting up individual and more complex studies in virtual environments. This toolkit prototype contributes to enhancing researchers capabilities in conducting social/behavioral studies in virtual worlds. The toolkit hopefully will empower social and behavioral researchers by proving a toolkit prototype that requires less time, efforts and costs to setup stimulus responses types of human subject studies in virtual worlds.

In the future, we plan to perform more extensive development of toolkit and conduct a more extensive study with more participants after more capabilities are developed. More future implementations would consist of including more advanced features for the toolkit such as including adaptive types of task (the answer to the previous question effects the next question), response scales and the ability to respond to the tasks with more than just yes and no. We will conduct a direct comparison between this toolkit and other similar applications. These features will provide more capabilities for the toolkit to setup and conduct more complex studies in virtual environments rather than limited to stimulus responses types of studies.

Acknowledgements. Thank you to participants of this research and who provided valuable feedback.

References

1. Friedman, D., Steed, A., Slater, M.: Spatial social behavior in second life. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 252–263. Springer, Heidelberg (2007)
2. Gregory, B.E.A.: How are australian higher education institutions contributing to change through innovative teaching and learning in virtual worlds? in changing demands, changing directions. In: Ascilite Hobart, pp. 475–490 (2011)
3. B. J..B. A. C. Loomis, J. M.: Virtual environment technology as a basic research tool in psychology. Behavior Research Methods, Instruments, and Computers 31, 557–564 (1999)
4. B. R.-M. D. Sanders, G.S.: Distraction and social comparison as mediators of social facilitation effects. Experimental Social Psych. 14, 291–303 (1978)
5. Moschini, E.: The second life researcher toolkit: an exploration of in world tools, methods and approaches for researching educational projects in second life. Springer (2010)
6. Antonio Gonzlez-Pardo, E. P., de Borja Rodriguez Ortz, F., Fernandez, D. C.: Using virtual worlds for behaviour clustering-based analysis. In 2010 ACM Workshop on Surreal Media and Virtual Cloning, pp. 9–14 (2010)
7. Lim, J.K.S., Edirisinghe, E.M.: Teaching computer science using second life as a learning environment. In: ICT: Providing Choices for Learners and Learning, p. 978 (2007)
8. Hayes, A.L., Ulinski, A.C., Hodges, L.F.: That Avatar is looking at Me! Social In-hibition in Virtual Worlds. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 454–467. Springer, Heidelberg (2010)
9. M. N. H. S. Zhang, Q.: A case study of communication and social interactions in learning in second life. In: 43rd Hawaii International Conference on System Sciences, p. 19 (2010)
10. Shailey Minocha, M.T., Reeves, A.J.: Conducting empirical research in virtual worlds: Experiences from two projects in second life. Virtual Worlds Research, vol. 3
11. Nick Yee, M.U.F.C., Bailenson, J.N., Merget, D.: The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. CyberPsychology and Behavior 18, 115–121 (2007)
12. S. E., Wiecha J., Heyden R, M. M.: Learning in a virtual world: experience with using second life for medical education. J. Med. Internet Res. 12 (2010)
13. Zanbaka, C., U. A.,G. P., L. F. H.: Effects of virtual human presence on task performance. In: International Conference on Artificial Reality and Telexistence (ICAT) (2004)
14. Zanbaka, C., Ulinski, A., Goolkasian, P., Hodges, L.F.: Social responses to virtual humans: Implications for future interface design, pp. 1561–1570. ACM Press (2007)