Eye-Tracking-based Wizard-of-Oz Usability Evaluation of an Emotional Display Agent Integrated to a Virtual Environment

Károly Hercegfi¹, Anita Komlódi², Máté Köles¹, Sarolta Tóvölgyi¹

¹Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Magyar tudósok körútja 2, 1117 Budapest, Hungary, hercegfi@erg.bme.hu, kolesm@erg.bme.hu, tovolgyi@erg.bme.hu

²Department of Information Systems, University of Maryland Baltimore County (UMBC), 1000 Hilltop Cir, Baltimore, MD 21250, USA, komlodi@umbc.edu

Abstract: This paper presents the results of the usability testing of an experimental component of the Virtual Collaboration Arena (VirCA) developed by the Cognitive Informatics Group of the Computer and Automation Research Institute of the Hungarian Academy of Sciences. This component is a semi-intelligent agent called the Emotional Display Object. We applied Wizard-of-Oz type high-fidelity early prototype evaluation technique to test the concept. The research focused on basic usability problems, and, in general, the perceptibility of the object as uncovered by eye-tracking and interview data; we analyzed and interpreted the results in correlation with the individual differences identified by a demographic questionnaire and psychological tests: the Myers-Briggs Type Indicator (MBTI), the Spatial-Visual Ability Paper Folding Test, and the Reading the Mind in the Eyes Test (RMET) – however, the main goal of this paper outreaches beyond the particular issues found and the development of an agent: it shows a case study on how complex concepts in Virtual Reality (VR) can be tested in very early stage of development.

Keywords: usability evaluation; Wizard-of-Oz; concept testing; early prototype; eye-tracking; human-robot interaction; virtual reality (VR); virtual agent; uncanny valley; individual differences

1 Introduction

The Emotional Display object as a virtual agent was integrated into a 3D virtual environment, the Virtual Collaboration Arena (VirCA) developed by the Cognitive Informatics Research Group of the Computer and Automation Research Institute of the Hungarian Academy of Sciences [1]–[3].
The research team of the Department of Ergonomics and Psychology at the Budapest University of Technology and Economics joined the project in its final phase to study the Emotional Display object in context of use.

1.1 Background of the Tested Emotional Display Agent: The Idea of Ethology-Inspired Robots

People do not like human-like robots, if the robots reach nearly perfect similarity with human robots, according to the “uncanny valley” hypothesis of Mori [4], [5].

Figure 1 shows that if a robot looks and works obviously as a machine, some similarity with humans can support its acceptance by humans in Human-Robot Interaction (HRI). Furthermore, for definitely machine-like robots, it looks true that the more the similarity with humans, the more affinity felt by humans.

![Figure 1](image-url)

The Uncanny Valley, as Mathur and Reichling [6] adapted the hypothesis from Mori [5], [4]

Theoretically, if a robot is indistinguishable from a human being, humans like to interact with it, as they would do with another real human. However, robots indistinguishable from human beings are parts of a possible future or they are featured in science-fiction novels and movies only, like some androids of the novel “Do Androids Dream of Electric Sheep?” by Philip K. Dick [7] (and the movie titled “Blade Runner” based on it). Until the realization of this – whether this is wanted or unwanted –, if the similarity of a human-like robot gets close to perfect similarity, humans feel uncanny instead of having a positive affinity: in this case, the robot looks to be like a moving corpse or zombie.

So, beside the developments in the direction of human-like humanoid robots [8]–[10], there are other directions to improve HRI. One promising way is the so-called etho-robotics: applying analogous ethological patterns borrowing not from the human-human interactions, but, for example, from the human-dog interactions. The domestication of dogs and more than 30,000 years of living humans and dogs
together resulted behavioural changes in dogs that enhanced the human-dog social interaction. This social competence of dogs may inspire a model for HRI [11]. Miklósi, Korondi, and their colleagues [12] argue that the robots’ embodiment and behaviour should fit their specific environments and functions: instead of aiming to build more and more human-like robots, various species can be considered as an analogue.

In this train of thought, we can go further: (1) Human-human interaction as a model for HRI can cause uncanny effect. (2) Human-dog interaction as a model for HRI can be applied, however, selecting the dog form the species can be considered to be arbitrary. (3) Interaction of humans and any existing species having social competence should be generalized. (4) Based on these experiences, social interaction of humans and new, non-human and non-animal type of artificial robots would come into the focus of the developers. One attempt to this generalization was developing a robot with wholly artificial form, but with social ability and behaviour that humans can respect [3], [13], [14]. It could mean a prospective realization of a new level of Human-Machine Interaction according to the Cognitive Infocommunications (CogInfoCom) approach [3], [13].

Researchers of the Department of Ethology at Eötvös Loránd University developed a virtual form consisting of a sphere and labyrinth-like perpendicular lines what can visualize emotions that participants of experiments can identify as emotional states of happiness, despair, fear and anger [13], [14]. So, it looks totally artificial. However, it inherited some behavioural ability from animal analogues to show emotions. (Referring the curve of Figure 1, this virtual “robot” can be placed close to 0 point of the horizontal dimension: it looks much more artificial than a humanoid robot, and looks more artificial than an industrial robot as well.) This always spinning Emotional Display object achieves these emotional states by changing its size, colour, and rotation rate, and pulsation of vertical position (“jumping”). This object is shown in Figure 2 and on the lower left side of Figure 3.

![Figure 2](image-url)

The Emotional Display object applied in this research. It is always spinning. Sometimes, it changes its size, colour, and rotation rate, or starts to pulsate its vertical position (“jumping”)
1.2 Related Works 1: Practice of Usability Evaluations in 3D Virtual Environments

In the research published in this paper, we aimed at performing the usability evaluation of a new object of a 3D virtual environment. The term ‘usability’ and the usability evaluation methods are matured in Human-Computer Interaction (HCI) [15], [16]. Already existing empirical usability evaluations, also in 3D virtual environments, have traditions: some of them focus on interaction styles, such as gesture control [17]; some of them focus on specific features of virtual environments, such as colours [18]; others focus on broad term of usability issues [19]–[24].

1.3 Related Works 2: Experiences on Individual Differences in Usability Evaluations

Usability evaluations often faltered at the level of identifying usability problems in general, without respecting the different needs of various users. However, the myth of “average user” [25] is exceeded by a number of studies involving individual differences, such as cultural background [26], personality types [27]–[29], and cognitive styles [30]–[32]. The current research published in this paper emphasizes the importance of the approach of taking care of individual differences.

2 Methods

2.1 Participants

Eighteen participants were involved, eleven female and seven male between the ages of 20-33 (mean age: 24).

2.2 Experimental Setting

All participants were seated in front of a Tobii T120 eye-tracking device, in the laboratory of the Department of Ergonomics and Psychology. Beyond eye-tracking, the user camera of the Tobii equipment also recorded the participants’ facial expressions.

The VirCA software and an intelligent cyber object of the VirCA system, the KUKA robot worked in the user’s computer.
The Emotional Display object and a user interface to control its “behaviour” were installed on an adjacent computer. The “behaviour” of the Emotional Display object was simulated by the experimenter following predefined rules and based on the user’s behaviour, as it is described below. This Wizard-of-Oz setup is often used in the evaluation of 3D environments, as described by Bowman et al. [33].

The virtual room of the VirCA contained a KUKA-type industrial robot, the tested Emotional Display object, some other objects, and some posters on the walls (Figure 3).

In mechanical engineering, the robot type called “KUKA” refers to a standard construction-type of industrial robots: a robotic arm, where the first axis is vertical, and then the next two axes are horizontal. The “KUKA” name came from the name of the German company that made it widespread in industry. On the margin, the applied virtual robot is a model of a particular, existing industrial robot (KUKA KR 6) made by the mentioned company. The participants were able to control this virtual industrial robot selecting it by a virtual cross-hair, then giving orders by a menu displayed in three dimensions.

Figure 3
The basic layout of the room in VirCA with the KUKA robot, two tables and three balls belonging to the KUKA object, an additional ball, and the tested Emotional Display object on the left side. The poster showing the instructions can also be seen in the background.
Around the KUKA robot, other related objects were placed: two tables and three balls. The menu of the KUKA contained a direct command to move a ball from one table to another. An additional (non-working, immobile) ball was placed on one of the tables of the KUKA.

The tested Emotional Display Object was also placed close to the industrial robot. It looked as it was described at the end of the Section 1.1 (Figure 2).

There were two additional objects (a domino and another ball) on the floor in the middle of the room. There were posters hanging on the walls. One of them showed the instructions for the participants.

2.3 Procedure

After the instructions, the participants completed three tasks with the VirCA:

1) The first task was to look at the posters on the walls at a glance just to practice a basic navigational task.

2) Then they had to delete two objects in the room using the menu to become more familiar with the object selection and menu functions.

3) Finally, they had to go to a given position and turn to a given direction to recreate the initial view they saw upon entering the space. This is how we ensured that the Emotional Display Object stayed in sight while the participant performed the actions. From this position, they had to command the industrial robot to move three balls from one table to another.

The overall position of the Emotional Display object was stationary all through the session, although it was not motionless. When it was idle, it had a continuous rotation; and, when it was in specific non-idle states, it changed its size, colour, and/or rotation speed as defined by its emotional state [13], [14]. The Emotional Display object had three emotional states thus behaving like an intelligent agent although operated by a human in the Wizard-of-Oz setup.

The predefined rules of the behaviour of the Emotional Display Object were the following:

• It showed the “happy” animation (animations were based on previous research [13], [14]) several times: first when the participant first saw the space, next when it was mentioned during the briefing, and then every time the participant succeeded in moving a ball.

• There were two other emotion animations, which were played in specific situations. If a participant failed to move a ball from one table to another in three tries, the Emotional Display Object displayed the “sad” animation. If someone failed to move a ball five times or made many wrong attempts at interacting with the system, it showed the “angry” animation.
As they were highly situational, most participants did not have the chance to see them. After the participants finished these tasks, they were interviewed about their experience with the system and completed several questionnaires:

- demographic questionnaire on their experience regarding First-Person Shooter and Simulator games or any other software that requires manipulating 3D environments (AutoCAD, etc.) and their experience with pets (what kind of pets and for how long did they have them),
- the RMET (Baron-Cohen’s Reading the Mind from Eyes Test [34]),
- a standard Paper-Folding Test [35],
- and the Hungarian version MBTI (Myers-Briggs Type Indicator [36]).

Our hypotheses were the following. First, we expected the participants with better spatial skills and more experience with 3D manipulation to be better at detecting any changes happening to the Emotional Display Object. We also expected that participants with higher RMET values to be able to recognize emotions simulated by the Emotional Display more easily. We also hypothesized that the longer time the participant had pets in the past, the more accurate his/her thoughts will be regarding the emotional state and overall functionality of the Emotional Display.

2.4 Statistical Analysis

Because of the type of the variables and the small number of cases, the connections between the variables were tested by carefully selected methods:

- In cases of testing connections between scale and ordinal variables, Spearman’s ρ (rho) correlation coefficients were calculated.
- In cases of testing differences between values grouped by dichotomy variables (comparing distributions across groups using grouping variable with two discrete values), non-parametric Mann-Whitney U tests were applied.

Statistical analyses were performed using the IBM SPSS Statistics software, version 22.0.
3 Results

3.1 Perceiving the Changes of the State of the Emotional Display Object

As the post-experiment interviews and eye-tracking data revealed, most of the participants did not see any change in the state of the Emotional Display Object at all. All were aware of the Emotional Display Object’s presence, because we mentioned it in the briefing as the representation of the industrial robot. However, most of the participants did not care to look at it during the tasks. This was somewhat expected based on earlier research [37]–[40]. The effect of selective attention can be quite powerful. A new task can use up all of the available attentional resources. This makes the participants focus only on the most important parts of the screen (Figure 4).

![Eye-tracking heat map of a typical interaction. Note that fixations only appear at areas needed exclusively for the task](image)

The surprising fact was that even when people fixated on the Emotional Display for a long time (4-5 s, Figure 5) while it was animated, they still were not able to report any changes in its behaviour during the post-experiment interviews.
Figure 5
Fixating on the Emotional Display while its animating didn’t ensure that the participant noted any changes

We also found that visibility may have been an issue. In its idle state, the Emotional Display differed from its background in colour. However, as you can see in Figure 6, as it plays the “happy” animation, its colour almost matches that of the background. This clearly worsened the visibility of the Emotional Display Object in a state when we wanted participants to notice it.

Also, the way the “happy” animation is presented made it more difficult to spot the Emotional Display. With the start of the animation, the object’s inner wire structure shrunk in size and levitated towards the floor. This, on some occasions, made some of its lower section to sink into the floor.

Since users could not interpret the emotions of the Emotional Display Object, we analyzed their ability to simply perceive and recognize the changes in the Emotional Display Object in the space.
3.2 Factors Impacting the Observation of the Emotional Display Object

We found a significant correlation between the number of fixations on the Emotional Display being in its idle state and in its animated state (Spearman’s $\rho = 0.779$, $p = 0.000$). This result means that the changes in its movement, size, shape, and colour did not draw any attention. This finding suggests that if someone noticed the Emotional Display Object at all then he/she looked at it more often irrespective of the movements and changes in it, and thus had a better chance of seeing its animations.

According to the scores of the MBTI psychological test, participants characterized with Feeling instead of Thinking fixated significantly longer (Spearman’s $\rho = 0.622$, $p = 0.006$), and for more often (Spearman’s $\rho = 0.553$, $p = 0.017$) on the Emotional Display Object.

Participants characterized more by Intuition than by Sensation also looked at the Emotional Display Object more often (Spearman’s $\rho = 0.459$, $p = 0.055$).
3.3 Factors Impacting the Observation of the Emotional Display Object’s Changes

We found a significant correlation between participants’ ability to connect the changes of the Emotional Display Object’s state to a particular action or event and the length of owning pets (Spearman’s $\rho = 0.423$, $p = 0.040$, Figure 7).

![Figure 7](image)

If the participants are grouped by their ability to connect the changes of the Emotional Display object’s state to a particular action or event, significant correlation can be identified: the longer time one had pets, the better his/her ability is to recognize the Emotional Display object’s reactions. Spearman’s $\rho = 0.423$, $p = 0.040$

It seems that the longer a participant had pets, the better his/her ability was to recognize the Emotional Display Object’s reactions. This result is promising for the future development of the Emotional Display Object.

The results of the Spatial-Visual Paper Folding Test also correlated strongly (Spearman’s $\rho = 0.638$, $p = 0.004$) with recognizing changes in the Emotional Display Object. The better the test result was, the more likely the participant was to attribute the observed change to some other action.

3.4 A Sample Usability Problem

A frequently occurring usability problem was passing the pointer hand through the object. This prevented the pointer from selecting the object.
According to the scores of the Myers-Briggs Type Indicator (MBTI) psychological test, participants characterized by Judgment instead of Perception were more likely to make this mistake as a trend ($p = 0.022$).

We identified a significant connection between the occurrence of this problem and a cognitive style dimension of the MBTI test: participants characterized by Judgment instead of Perception were more likely to make this mistake as a trend (Mann-Whitney’s $U$ test, $p = 0.022$, Figure 8).

### 3.5 Confusion between the Role of the Emotional Display Object and the KUKA Industrial Robot

Another frequently occurring problem was that some of the users selected the Emotional Display Object instead of the KUKA industrial robot to command the robot. This happened in spite of the participants being instructed: “This rotating structure is a small creature who is a representation of the internal state of the robot.”

We identified a tendency-like connection between the occurrence of this problem and a cognitive style dimension of the MBTI test: participants characterized by Thinking instead of Feeling were more likely to make this mistake as a trend (Mann-Whitney’s $U$ test, $p = 0.088$, Figure 9).

Those who avoided this problem were more likely to notice changes in the Emotional Display object (Mann-Whitney’s $U$ test, $p = 0.019$).
4 Discussion and Suggestions

The users were too focused on their tasks to notice the Emotional Display Object. A possible explanation of this is that the users in our experiment were all inexperienced in the use of the environment and did not have a chance to spend longer practicing the interaction methods. As previously demonstrated in the literature [37]–[40], selective attention can be a powerful effect in given circumstances. In longer and/or multiple interaction sessions users could gain significantly more experience with the interaction. This experience would reduce their cognitive load resulting from the interaction and better allow them to observe and recognize the Emotional Display Object. Thus, a study with multiple interaction sessions may result in different perceptions.

The Emotional Display Object’s lack of movement may have also contributed to its “invisibility” to some of the participants. Our original idea included a moving agent that could have guided the gaze of the participants by flying closer to objects relevant to the task at hand.

This dynamic location changing behaviour of the Emotional Display Object could make it more visible and functional. Its motion alone would draw more attention to it while directing the gaze of the participant to the objects required to finish a
task. In its current, stationary form, it often leaves the users field of view. By dynamically adjusting to the field of vision (by constantly staying in the lower left or right corners of the screen), it would assure its visibility. Beyond that it would not seem like just another object in the virtual space but it could be a part of the user interface. This, however, could not have been implemented by the time of our data collection.

The development process can continue in a new VR environment: the successor of the VirCA system applied now is the MaxWhere VR platform [41]–[44] that promises new prospects.

On the grounds of our research, further ideas of development can be produced that can later be evaluated in a set of studies that would also incorporate the lessons learned for the evaluation of 3D spaces. The lessons learned from this research have outreached beyond the particular issues found and the development of an intelligent object: this case study have shown a successful practice of methods capable of testing complex concepts in Virtual Reality (VR) in very early stage of development.

Acknowledgement

The authors thank Péter Baranyi, Péter Korondi, István Fülöp, György Persa, Ádám Miklósi, and Márdta Gácsi for the exciting cooperation, and Eszter Józsa, Gyöngyi Rózsa, and Judit Boross for their research assistance and preparation of the publication. This research was supported by the ETOCOM project through the Hungarian National Development Agency in the framework of Social Renewal Operative Programme supported by EU and co-financed by the European Social Fund.

References


