Perception of Spatial Relations and of Coexistence with Virtual Agents

Mohammad Obaid[1], Radosław Niewiadomski[2], and Catherine Pelachaud[2]

[1] Human Interface Technology Lab New Zealand (HITLab NZ), University of Canterbury, Christchurch, New Zealand
mohammad.obaid@hitlabnz.org
[2] CNRS-LTCI, Telecom ParisTech, Paris, France
{catherine.pelachaud,niewiado}@telecom-paristech.fr

Abstract. This paper focuses on the user's perception of virtual agents embedded in real and virtual worlds. In particular, we analyze the perception of spatial relations and the perception of coexistence. For this purpose, we measure the user's voice compensation which is one of the human automatic behaviors to their surrounding environment.

The results of our evaluation study reveal that people compensate their voice according to the distance during the interaction with both augmented reality (AR) and virtual reality (VR) based agents. Secondly, in AR-based scenario users perceive stronger the distance between them and the virtual agent. On the other hand, the results do not show any significant differences regarding the notion of coexistence of the user. Finally, we discuss our results in the context of sense of presence in interaction with virtual agent in AR applications.

Keywords: Augmented Reality, Immersive Virtual Reality, Virtual Agents.

1 Introduction

In this paper we address two aspects of the user's perceptual experience when interacting with expressive Augmented Reality (AR)-based agents. In particular, we look at (1) the user's perception of spatial relations with an agent, and (2) the user's perception of coexistence within the environment of an agent. Both aspects directly refer to the notions of "presence". For this purpose we conducted a scenario based user evaluation study that compares the user's responses to an agent in AR and Virtual Reality (VR) environments. The scenario is based on the fact that in human-human communication [1] people speak louder when they believe that they are more distant from each other. Therefore, we expect that the user's reaction is similar in human-agent communication.

When evaluating interaction though different media, such as AR and VR, often the concept of "sense of presence" is used. This concept is usually understood as "subjective experience of being in one place, even when one is physically situated in another" [2]. Thus, it is a subjective state or feeling which includes the

H. Högni Vilhjálmsson et al. (Eds.): IVA 2011, LNAI 6895, pp. 363–369, 2011.

notion of “being there” (for VR environments) or “coexistence in the same space” (for AR environments)[1]. According to Slater et al. [3], a high sense of presence occurs when humans respond to computer-generated data as if they were real. Various forms of human reactions can be then observed from a low level psychological processes (e.g. arousal), by unconscious automatic behaviors (e.g. avoid obstacles) and reflexes, to high-level voluntary (and conscious) behaviors (e.g. speech acts). This extended definition of the sense of presence permits to not only measure subjective impression by post-experiment questionnaires but also to measure more objectively human responses to different environments. Several researchers used behavioral measurements such as skin conductivity [4] or postural responses [5]. In our work we propose to use a Sound Pressure Level (SPL) meter to measure the unconscious adaptation of the volume of the voice when addressing an interlocutor at a long distance. According to Warren [1], people implicitly use “tacit knowledge of the sensory effects of changing distance” and adjust their voice volume based on personal experiences when communicating with distant listeners. Finally, people compensate their voice volume in interaction with a physical speaker but also with an imagined one [6]. Consequently, we expect that this compensation effect will also occur in the interaction with the virtual agent and that people will react to computer-generated data in AR/VR similarly to human-human interaction. We also wonder if embedding the agent in the real world (as in AR) or in the virtual world (as in VR) has an effect on voice compensation.

On the other hand, presenting a virtual agent in an AR environment to the user is a challenging task. In a VR environment, agents are perceived by the user as digital content just as the rest of the surrounding environment. However, embedding virtual agents within a real environment can introduce some visual incoherency to users due to lighting, shadows or occlusions. Consequently, one may expect (e.g. [7,8]) that perception of the coexistence and interaction in AR-based environments can be worse than the classical VR-based one.

2 Experimental Evaluation

Our experiment extends the previous research (e.g [8,7]) by evaluating the perception of the distance of the virtual agent in AR and VR and the user’s perception of coexistence in the same environment as of the virtual agent. The hypotheses of the designed experiment are: (H1) The voice compensation for the change in distance occurs in both AR and VR. (H2) Spatial relations and the distance are stronger perceived by the user during the interaction with an AR-based agent. (H3) The perception of coexistence is higher during the interaction with a VR-based agent.

2.1 Design

Scenario: Our scenario exploits spatial relations between the human user and the virtual agent during the interaction. Initially, the virtual agent is standing

[1] In the rest of the paper we use the term coexistence for both types of environments.

approximately 15 meters (Point A) from the user. As the agent is setup not to be facing the user, the task of the user is to attract the agent's attention by calling it. After few calls the agent turns around and re-appears in front of the user at a natural communication distance of approx. 1 meter (Point B). The user is then engaged with the agent by asking it to do a simple task (count to three).

Implementation: We used the Greta agent [9] for both conditions. In the AR version (1) the system displays Greta in the real space (see Figure 1b; for the details of the implementation see [9]). Greta initially is standing at the end of a corridor in position (Point A). The user uses a Head Mounted Display (HMD) to interact with Greta. In the VR version (2) the system shows Greta in an immersive 3D environment that is setup to be similar to the AR environment (see Figure 1a). A 3D model of the corridor used in the AR condition is used to simulate the same space as in the AR environment. Greta is initially standing at the end of virtual corridor. In both conditions Greta's body size ratios are all kept the same as much as possible.

2.2 Procedure

Each subject participated in only one out of the two conditions. At the beginning each participant was given a description of their tasks and was told that the character's name that they will be communicating with is Greta, but was not told how loud they need to speak with Greta. Each subject was also asked about his/her level of expertise with the AR (or VR accordingly) and, based on their answer, a short demo was given to participants who are not familiar with the AR (or the VR accordingly) environments. The demo was not related to the content of the experiment.

In the AR condition the procedure starts by positioning the user in the corridor and asking them to wear the HMD. In the VR condition the procedure starts by positioning the user 1.5 meters in front of a three-screen back projection system[2]. Then they were told to put the 3D glasses and headphones on before starting the Greta-corridor 3D scene.

Fig. 1. (a) Greta in the VR condition. (b) Greta in the AR condition.

[2] The virtual environment was displayed using a three-screen back projection system with a field of view of 180 degrees. Each screen measured 2.44m x 1.83m.

In both conditions users were told to start communicating with Greta approximately 5 seconds after she appears on their display. This is mainly to get them familiar with the HMD (resp. 3D screen) setup before they start communicating with Greta. The time is also used by the experiment's administrator to start the SPL meter. Users were not aware that there was a SPL meter device near them. They were not allowed to move from their position when interacting with Greta. We used the Wizard-of-Oz approach to operate Greta.

Behavior Measurement: To assess the user's loudness level in both conditions we used a professional Sound Pressure Level meter (SPL) application for iPhone-4. We recorded the peak SPL measurement using A-weighting. Sound pressure level is measured in decibels, dB to the standard reference level of 20μPa (this is the threshold of human hearing at 1kHz). The SPL meter application comes pre-calibrated for the typical iPhone-4 built-in headset mic to +7dB. The SPL meter is placed approximately 1 meter away from the user. The measurements of the user's voice were done for two localizations of the agent in both conditions (AR/VR) i.e. at *point A* - when the user tries to call (measurement SPL_A) and at *point B* - when the user asks her to count (measurement SPL_B).

Questionnaire: After the experiment is finished, participants were asked to complete a questionnaire that contains fourteen questions. Twelve questions have a scale from 1-5, where 1 mean strongly disagree and 5 means strongly agree: (Q1) I felt that Greta was really down the corridor. (Q2) I felt as though I was in the same space as Greta. (Q3) I would have liked the experience to continue. (Q4) I felt Greta heard me. (Q6) I felt Greta responded to (resp. understood) my call. (Q7) I felt Greta wanted to start a conversation with me. (Q8) Did you think you had to shout to call for Greta? (Q9) Did you shout to call for Greta? (Q10) Did you feel getting closer to Greta before calling her? (Q11) (resp. (Q12) I felt Greta was far from me in point A (resp. point B). Additionally, two questions (Q13 and Q14) asked the user to estimate how far Greta was in meters at the two points A and B. Users were asked to choose one answer from five: *1m, 3m, 5m, 9m, 15m.* 32 subjects participated in the study. All participants are from New Zealand and are aged between 20 and 63 years old (mean 29.7, SD=9.1). 12 males and 4 females participated in the AR condition with an average age of approximately 32 years. 7 participants have not had any experience with AR in the past. 12 and 4 females participated in the VR condition with an average age of approximately 28 years. 8 participants have not experienced being in an immersive 3D environment in the past.

3 Results

For hypothesis H1, we calculated separately (for each of the conditions AR/VR) the difference between the voice volume at point A and B as well as the difference in perceived distances of the agent (questions Q11-Q14). Table 1 outlines the captured values. In the AR condition, the SPL at point A varied from 54.8 dB to 66.1 dB with a mean value of 60.96 dB, while the SPL at point B varied from

Table 1. The mean values of the users' answers for questions Q11 - Q14 and SPL measurements at point A and B. Standard deviations appear in parentheses.

Condition	SPL_A	SPL_B	Q11	Q12	Q13	Q14
AR	60.96 (3.71)	48.33 (2.70)	4.31 (0.70)	1.5 (0.63)	14.25 (2.05)	1.13 (0.5)
VR	57.92 (4.00)	51.46 (3.66)	4.13 (0.72)	1.69 (0.60	11.25 (3)	2.38 (1.20)

Table 2. The mean values of the users' answers for questions Q1 - Q10. Standard deviations appear in parentheses. Significant differences in bold.

	Environment		Expertise	
	AR	VR	no	yes
Q1	4.31 (0.70)	4.13 (0.72)	3.93 (0.70)	**4.47** (0.62)
Q2	3.56 (0.96)	3.81 (0.98)	3.27 (0.80)	**4.06** (0.97)
Q3	3.13 (1.02)	3.5 (0.97)	2.80 (1.01)	**3.76** (0.76)
Q4	3.88 (1.02)	**4.44** (0.63)	3.93 (1.03)	4.35 (0.70)
Q5	4.06 (0.93)	4.44 (0.63)	3.87 (0.92)	**4.59** (0.51)
Q6	3.81 (0.91)	**4.25** (0.77)	3.53 (0.83)	**4.47** (0.62)
Q7	2.5 (1.21)	2.31 (0.79)	2.47 (0.74)	2.35 (1.22)
Q8	**2.81** (1.11)	2 (0.73)	2.13 (0.83)	2.63 (1.11)
Q9	**4** (0.82)	2.38 (0.81)	2.67 (0.90)	**3.65** (1.17)
Q10	**2.13** (0.81)	1.56 (0.63)	1.60 (0.90)	3.65 (1.17)

42.8 dB to 53.8 dB with a mean of 48.33 dB. The maximum difference observed for one subject was 17.5 dB and the mean difference was 12.63 dB. The repeated-measure ANOVA test revealed that the SPL difference between points A and B is significant ($F(1,15) = 213$, $p < 0.0001$). Moreover, the post-study questionnaire results show that the mean of the perceived distance between points A and B (obtained from questions Q13 and Q14) is 13.13 meters, while difference between the mean answers for Q11 and Q12 is 2.81 out 5 points in Likert scale. In the VR condition, the SPL at point A varied from 50.2 dB to 63.7 dB with a mean value of 57.92 dB. The SPL at point B varied from 44.3 dB to 57.6 dB with a mean of 51.46 dB. The maximum observed difference was 11.4 dB and the mean difference was 6.46 dB. The repeated-measure ANOVA test revealed that the SPL difference between points A and B is significant ($F(1,15) = 80$, $p < 0.0001$). In the case of the post-study questionnaire, the perceived mean distance between points A and B (Q13 - Q14) is 8.88 meters, while the mean difference between the answers for Q11 and Q12 is 2.44 out 5 points in Likert scale.

Secondly, to check the effect of the Environment variable and the users' Expertise level variable we conducted a 2×2 ($Environment \times Expertise$) between subject MANOVA on questions Q1 - Q10. Results show two main effects, one of the Environment [$F(12,17) = 5.147$, Wilks' $\lambda = 0.216$, $p = 0.001$] and one of the Expertise [$F(12,17) = 4.609$, Wilks' $\lambda = 0.235$, $p = 0.002$] and no interaction effect [$F(12,17) = 0.846$, Wilks' $\lambda = 0.625$, $p = 0.609$]. Distinguishing between independent variables, results show an effect of Environment for questions Q4 ($F(1,31) = 4.349$, $p = 0.046$) , Q6 ($F(1,31) = 4.280$, $p = 0.048$), Q8 ($F(1,31) =$

5.875, p = 0.022), Q9 (F(1,31) = 42.813, p=0.0001) and Q10 (F(1,31) = 4.432, p = 0.044). Moreover, an effect of the Expertise was observed on the questions Q1, Q2, Q3, Q5, Q6 and Q9. Table 2 shows the detailed results.

To detect the most important factors we performed PCA on Q1 - Q10, which resulted in two factors. The first factor (GR1) regrouped questions Q1 - Q6. 5 out these 6 appear to be related to the concept of the coexistence. The second factor (GR2) regroups Q7 - Q10. 3 of them (Q8 - Q10) appear to be related to the perception of distance. Question Q7 has the most balanced values (0.485, 0.549) and cannot be easily classified to any of these two concepts. ANOVA test revealed an effect of condition (AR/VR) on group GR2 (F(1,31) = 10.841, p = 0.003). While, no effect was observed for GR1 (F=(1,31) = 1.606, p = 0.215).

4 Discussion

The aim is to study the perception of the distance and the coexistence, which are two very important issues for AR-based applications. According to the obtained results, we confirm hypothesis H1 (regarding the voice compensation) for both environments. The human automatic responses (SPL_A and SPL_B) to the agents located at points A and B are significantly different, while we find that this difference is stronger in the AR condition. The significant difference in the sound level measurements is also supported by the results of Q11 - Q14. Participants consciously perceive the difference in distance between the agent placed at point A and B. Interestingly, the results show that in the AR condition this difference is higher.

Hypothesis H2 focuses on the perception of distance and is addressed by Q8 - Q10 of the post-study questionnaire. The participants' results from those questions receive significantly higher results in AR. Consequently, we can say that hypothesis H2 is confirmed and people perceive the spatial relations stronger in AR.

The results corresponding to hypothesis H3 are ambiguous. We expected that the user's perception of coexistence might be lower in the AR condition. The results for the questions Q4 - "I felt Greta heard me" and Q6 - "I felt Greta understood my call" are lower in the AR condition. One may presume that incoherency between the embedded animation and real image may give the user the impression of not sharing the environment with the virtual agent and thus participants may have the impression that they do not exchange directly the messages with the virtual agent. On the other hand, no significant difference was observed for questions Q1 - "I felt that Greta was really down the corridor" and Q2 - "I felt as though I was in the same space as Greta". Finally, we do not observe any significant difference in the GR1 regrouping questions Q1 - Q6. This makes us think that the "perception of coexistence" with the virtual agent is not lower in the AR environment.

5 Conclusion

In this paper we presented a comparative study of virtual agents in two environments, AR and VR. In the study we used an objective measure of the user's voice loudness level, which is one of the human automatic behaviors to their surrounding environment. Firstly, the results of the evaluation revealed that humans compensate their voice level during the interaction with both AR and VR-based agents. Secondly, users strongly perceive the distance between them and the AR-based virtual agent. On the other hand, the results did not show any significant improvements to the perception of coexistence of the user in the space of the AR-based agent compared with the VR-based agent.

The other interesting conclusion is that we are able to confirm the subjective user impressions by measuring objectively the user's voice loudness level. People use similar mechanism of voice compensation according to the interlocutor's distance when interacting with AR or VR-based agent as when interacting with other human. Voice compensation observed in our experiment is one of the humans' automatic responses towards a virtual entity that do not have any rational motivation as it was already pointed out by several studies [10]. We confirm that people interacting with the virtual agent in AR/VR tend to re-use unconsciously their "natural" behavior patterns learned in "real world" interactions.

References

1. Warren, R.M.: Measurement of sensory intensity. Behavioral and Brain Sciences, 175–189 (1981)
2. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: A presence questionnaire. Presence: Teleoper. Virtual Environ. 7, 225–240 (1998)
3. Slater, M., Lotto, B., Arnold, M.M., Sanchez-Vives, M.V.: How we experience immersive virtual environments: the concept of presence and its measurement. Anuario de Psicologa 40, 193–210 (2009)
4. Llobera, J., Spanlang, B., Ruffini, G., Slater, M.: Proxemics with multiple dynamic characters in an immersive virtual environment. ACM Trans. Appl. Percept. 8, 3:1–3:12 (2010)
5. Freeman, J., Avons, S., Meddis, R., Pearson, D.E., Ijsselsteijn, W.: Using behavioral realism to estimate presence: A study of the utility of postural responses to motion stimuli. Presence: Teleoperators and Virtual Environments 9(2), 149–164 (2000)
6. Healey, E.C., Jones, R., Berky, R.: Effects of perceived listeners on speakers? vocal intensity. Journal of Voice: Official Journal of the Voice Foundation 11(1), 67–73 (1997)
7. Dow, S., Mehta, M., Harmon, E., MacIntyre, B., Mateas, M.: Presence and engagement in an interactive drama. In: Proceedings of the SIGCHI Conference CHI 2007, pp. 1475–1484. ACM, New York (2007)
8. Tang, A., Biocca, F., Lim, L.: Comparing differences in presence during social interaction in augmented reality versus virtual reality environments: An exploratory study. In: Raya, M.A., Solaz, B.R. (eds.) Proceedings of PRESENCE 2004, 7th Annual International Workshop on Presence, pp. 204–208 (2004)
9. Niewiadomski, R., Bevacqua, E., Le, Q.A., Obaid, M., Looser, J., Pelachaud, C.: Cross media agent platform. To appear in: Proceedings of Web3D ACM Conference, ACM, New York (2011)
10. Reeves, B., Nass, C.: The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York (1996)