“Hold My Hand, Baby”: Understanding Engagement through the Illusion of Touch between Human and Agent

David Jan Mercado
LTCI, CNRS, Télécom ParisTech,
Université Paris-Saclay
Paris, France
davidjanmercado@acm.org

Gilles Bailly
LTCI, CNRS, Télécom ParisTech,
Université Paris-Saclay
Paris, France
gilles.bailly@telecom-paristech.fr

Catherine Pelachaud
LTCI, CNRS, Télécom ParisTech,
Université Paris-Saclay
Paris, France
pelachaud@telecom-paristech.fr

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
CHI'16 Extended Abstracts, May 07-12, 2016, San Jose, CA, USA
ACM 978-1-4503-4082-3/16/05.
http://dx.doi.org/10.1145/2851581.2892463

Abstract
This paper explores the quality of engagement between human and agent in large displays. We designed a musical application and introduced the *illusion of touch* as a novel interaction concept where the user and the agent interact through mediated virtual touch. We conducted a lab study to better understand the users’ behaviors and engagement with our system. Our main findings from the design exploration are: 1) users are engaged in the interaction with an interactive agent even though the agent is not perceived as engaged by the user; 2) users are more involved in the experience when the agent is interactive; and 3) the agent’s gaze is effective in grabbing the user’s attention. These findings provide initial insights on designing for engagement in displays that make use of embodied agents and should be validated with a field study.

Author Keywords
HCI; Engagement; Embodied Agents; Large Display; Illusion of Touch

ACM Classification Keywords
H.5.1. Multimedia Information Systems: Artificial, augmented, and virtual realities

Figure 1: *Illusion of touch*, a novel interaction concept where the user and the agent interact through mediated virtual touch.

Figure 2: Greta is a three-dimensional embodied agent used in a number of European projects such as CALLAS and SEMAINE.

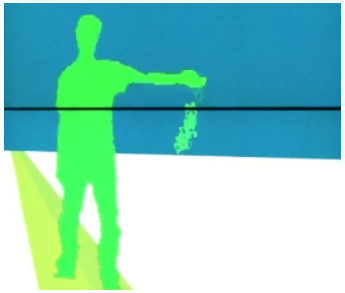

Figure 3: Musical application with piano keys on the virtual floor depicted in shades of green for each note while notes trailed after the user's hand.

Introduction

Public displays have been used to foster social interaction between and among people in recent years [8, 1, 2]. The use of embodied agents in public displays, however, has not been investigated yet within the context of user engagement. Our study recognizes the ability of an agent to gaze as the necessary first step to engagement and thus makes use of this capacity to grab attention. It goes further by giving the agent the ability to respond to the user's touch.

The direct sensation of touch between humans triggers physical connection at the base level and emotional attachment on a deeper level. Direct touch is difficult to establish between human and agent, however. Multi-touch screens and haptic devices can simulate this in virtual environments; but when hygiene and multiple interactions are concerned, it is impractical, if not impossible, to deploy such a complex system in public. This paper argues that through the *illusion of touch*, humans are more engaged in the interaction with an agent. We introduced this concept because users are engaged in (virtual) touch interaction while there is no physical contact involved (Figure 1). This engagement provides a novel sense of connection between human and agent that has the potential to facilitate and encourage social interaction.

Engagement is defined as the "value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction" [10]. This definition has been used in several studies measuring human-agent engagement [9, 5]. Essential to this notion is the cognitive and emotional involvement and commitment that should be established between users and agents. An agent therefore must be capable of perceiving, adapting to, and generating behaviors relating to attention and emotion [9].

We examined how the *illusion of touch* affects engagement in human-agent interaction. The musical application we designed showed an agent in an empty scene with piano keys on the virtual floor for playing music that was meant to elicit curiosity and encourage exploration.

Related Work

Our research is at the crossroads of Embodied Conversational Agents (ECAs) and Human-Computer Interaction (HCI). ECAs are "conversational in their behaviors and specifically human-like in the way they use their bodies in conversation" [3]. They are able to express and recognize both verbal and nonverbal behaviors and have the ability to respond to users through body movements and facial expressions.

Research in ECA explores a number of topics including emotion [7] and engagement [13, 9]. Its use in public displays has not been fully explored yet. A research on intelligent kiosks found that an agent (an animated talking head) that speaks to and turns its gaze towards the user is effective in attracting attention and showing a sense of awareness [4]. Within the context of the shopping experience, Reitberger et al. [12] developed the Persuasive Interactive Mannequin (PIM), a virtual representation of a real mannequin, that has two goals: to grab customer attention and to increase the time spent in front of a shop window. It is found that the reaction and recognition of the user's presence is effective in persuasion at the microsuasion level where incremental behavior and attitude changes are expected.

Figure 4: Non-interactive condition where the agent shows minimal nonverbal behaviors. The agent looks around the virtual scene while the user explores the application.

Figure 5: Interactive condition where the agent gazes towards the user as the agent responds to the user's *touch*.

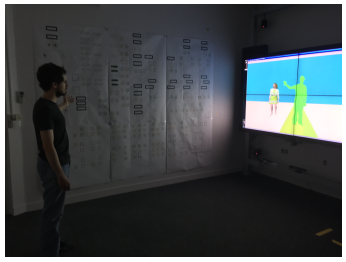

Figure 6: Participant exploring the musical application.

System Design

The musical application uses Unity 3D to display 3D content (Figure 3), Microsoft Kinect and OpenNI to track users, and Greta to serve as the 3D embodied agent (Figure 2). The conversational aspect of the agent was not used in this study; hence, we used the term embodied agent or agent throughout this paper. The agent's animations were downloaded from Carnegie-Mellon University's motion capture library. The agent was idle (looking around the virtual scene) when there was no one present but gazed towards the participant and responded to its touch otherwise (Figure 1). A virtual piano was included to add dynamism in playing music (Figure 3). The instrumental of the children's song "If you're happy and you know it" was chosen to motivate interaction, the rhythm of which blended well with the virtual piano. The notes that trailed after the user's hands visually hinted at volume mapping – the farther the right and left hands were to each other, the louder the music became. The user's image was shown as a silhouette on the screen as it afforded anonymity and stimulated artistic freedom [8]. In the Interactive condition described in the next section, the lyrics started to play whenever the user's hand *touched* the agent's hand, which gave the impression that the agent was singing along.

Experimental Design

The objective of the lab study was to get insights on user behavior and to explore how the *illusion of touch* engages humans in an interaction with the agent.

Conditions. There were two experimental conditions: (A) With Non-interactive Embodied Agent (Non-interactive condition) and (B) With Interactive Embodied Agent (Interactive condition). A Non-interactive agent shows minimal nonverbal behaviors such as gazing and looking around the virtual scene but it cannot directly interact with the user (Figure 4) whereas an Interactive agent has the ability to interact with the user by *holding* its hand (Figure 5). Interaction of the agent with the user is strictly defined in our research as the mediated virtual touch where the user's mirrored silhouette *touches* the agent's hand (Figure 1).

Participants. Ten students (mean age = 25.3; 9 men and 1 woman) participated in a within-subject study.

Task and procedure. The task was designed to simulate public displays where first-time users did not know what the application was about until they started interacting with it. Participants answered a pre-questionnaire. They were then tasked to freely explore the system deployed in a large screen display (2.5m x 1.5m) without time limit and with no knowledge of the interaction possibilities (Figure 6). Similar instructions were given in both conditions. Each session ended whenever the participant decided to finish the exploration. Finally, they answered a post-questionnaire after each condition and were interviewed briefly after completing both conditions.

Questionnaires. The pre-questionnaire asked for the participants' basic information and their familiarity with embodied agents and virtual environment while the post-questionnaire was grounded on Poggi's definition of engagement [10] and was selected from the items in the Temple Presence Inventory (TPI) [6] relevant to the features we considered in our engagement research.

Design. Half of the participants were treated with the Non-interactive condition first followed by the Interactive condition while the other half was treated with the reverse to counterbalance the order.

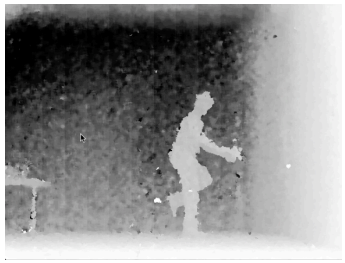

Figure 7: Participant attempted to dance with the agent.

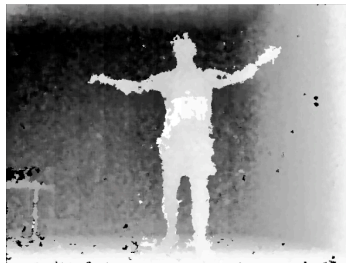

Figure 8: Participant performed wide arm gestures.

Figure 9: Participant bent to *hold* the agent's hand.

Data Analysis

Survey Data

Two-tailed paired t-tests were used to analyze items in the questionnaire that were asked in both conditions while one sample t-tests with hypothetical mean of 4 (midpoint of the 7-point Likert scale) were used otherwise. One outlier was removed after confirming it was significant ($p<0.01$) based on Grubbs' test and after careful video analysis that the participant did not do the action the specific item in question asked for ($n = 9$).

Based on Poggi's definition of engagement, the participants felt engaged in the interaction with the agent (Table 1). They reported highly that they both wanted to interact and wanted to continue the interaction with the agent in the Interactive condition; thereby affirming the two components of engagement pointed out by Poggi [10]. In the Non-interactive condition where holding the agent's hand was not possible, the participants reported highly that they tried to interact with the agent even though no reaction took place. From the agent's perspective, however, the users did not feel that the agent was engaged primarily due to its reactive nature as opposed to it proactively seeking for interaction.

Based on the Temple Presence Inventory, the participants were completely involved in the experience but they were not mentally immersed nor were their senses engaged (Table 2, 2.1-2.4). The participants in the Interactive condition also felt they could reach and hold the agent's hand while those in the Non-interactive condition reported they did not feel that way (Table 2, 2.5-2.6).

The gaze of the agent was observed in both conditions (albeit noticing it more in the Non-interactive condition), which confirms past studies on displays that successfully grabbed the attention of passers-by through gaze [4, 12]. Since this cue was present in both conditions, it cannot be the sole factor why engagement was significant (Table 2, 2.7-2.8). Nevertheless, this confirms that attention in the form of gaze is a precursor to engagement – in the absence of which engagement may not exist [9].

The participants reported wanting to and trying to make eye contact with the agent despite not having the sensation that the agent could see them, which they did so more often in the Non-interactive condition (Table 2, 2.11-2.14). This is consistent with the findings that the participants noticed the absence of eye contact with the agent more in the Non-interactive condition.

Video Data

Roughly sixty-five minutes of depth map video logs were analyzed to identify patterns of behavior. The most surprising observation was to see the participants dance and encourage the agent to dance with them as well (Figure 7) while the most common hand gesture was spreading the arms sideways while raising it up and down (Figure 8). These behaviors were similar to the findings of Tomitsch and his colleagues [14]. The playful gestures can make the experience in displays richer. Unlike previous studies where only half of the user's silhouette was shown [8, 16], our study showed the full body contour of the user, which allowed the participants to use both their hands and feet to interact with the system. When trying to hold the agent's hand, most participants started with a handshake pose and held it still for a while (Figure 9).

A number of participants talked directly to the agent even though the agent was not endowed with conversational abilities. One participant called the agent with endearment: "Hold my hand, baby". This confirms that

computers, more specifically agents, are treated the same way humans would treat a fellow human [11].

Implications for Human-Agent Interaction in Large Displays

Users are most likely to explore the space initially by foot based on our observations. Gestures are secondary but not necessarily less important. In designing displays in spacious settings, freedom of movement is necessary to allow maximum discoverability of the system's features. Hand and feet interaction whether in the form of dance or other full-body movements are also seen as playful. Future studies looking closely at this could map these fluid movements to specific meanings that can trigger actions to create a more intuitive and richer experience.

Establishing awareness that an agent gazes at humans is an essential first step to building and eventually sustaining engagement. Our findings show that the agent's ability to gaze towards the user's silhouette has been observed as the point where engagement begins.

The ability of simulating an *illusion of touch* through the user's silhouette presents a novel way of interaction between human and agent in large displays. It is a potentially powerful concept that captures the need for a mediated virtual touch in interaction. The agent's behaviors should extend beyond the screen and *reach* out to develop connection and possibly to express more complex emotions not only through facial and vocal behaviors but also through bodily contact.

Van Erp and Toet [15] made the assertion that the communication of emotions and effectiveness of social agents and robots will increase by providing them the ability to touch. Our study is a first attempt to build on what the authors had laid out in its guidelines. The concept we introduced in this paper provides a ground for subsequent research in interpersonal communication between human and agent, which at the very least, begins by ensuring that engagement exists.

Our study can be applied to other domains as well. When agents are used as artificial companions, the *illusion of touch* can add another layer in the communication pipeline. The use of this can also make for a fuller and more enjoyable experience in immersive gaming.

Future Work

This research provides initial insights to how the *illusion of touch* is a promising interaction concept to understand engagement between human and agent in large displays. The lab study has shown that from the user's perspective, interaction with the interactive agent resulted to better engagement even though the agent is not perceived as engaged by the user.

We recognize the current limitations of our work and the need to conduct field study to validate the initial findings. This research opens up new directions to understand how the *illusion* as opposed to the *sensation* of touch can engage human and agent. The comparison between the two could pave way for research that explores whether perception outweighs sensation in human-agent interaction. Our definition of interaction has been very limited as well. It is necessary to investigate other ways where interaction also involves the agent's conversational aspect and the nuances in communication on top of its ability to respond to touch.

Questions	M_{diff}	SD_{Mdiff}	t
1.1 To what extent did you want to interact with the agent?*	-1.78	1.72	3.11
1.2 To what extent did you try to interact with the agent? (Condition A)*	1.33	NA	3.27
1.2 To what extent did you want to continue the interaction with the agent? (Condition B)**	1.11	NA	5.55
1.3 To what extent did you think the agent would like to interact with you?	-0.89	2.20	1.21
1.4 To what extent did you think the agent tried to interact with you? (Condition A)**	-1.89	NA	4.86
1.4 To what extent did you think the agent would like to continue the interaction with you? (Condition B)	-1.00	NA	1.60

Table 1[1]: Questions based on Poggi's definition of engagement (*p<0.05; **p<0.01).

Questions	M_{diff}	SD_{Mdiff}	t
2.1 To what extent did you feel mentally immersed in the experience?	-1.00	1.58	1.90
2.2 How involving was the experience?*	-1.22	1.20	3.05
2.3 How completely were your senses engaged?	-0.78	1.39	1.67
2.4 How relaxing or exciting was the experience?	0.11	1.05	0.32
2.5 How often did you want to or try to hold the agent's hand?*	-2.11	2.47	2.56
2.6 How much did it seem as if you could reach out and hold the agent's hand?*	-2.00	2.18	2.75
2.7 How well were you able to observe the nonverbal behaviors of the agent?	0.11	1.69	0.20
2.8 How well were you able to observe the gaze of the agent?**	1.33	1.12	3.58
2.9 How often did you make a sound out loud (e.g. laugh or speak) in response to the agent?	-1.11	2.26	1.47
2.10 How often did you smile in response to the agent?	-0.56	1.42	1.17
2.11 How often did you have the sensation that the agent could also see you?	-0.11	1.17	0.29
2.12 How often did you want to or did you make eye contact with the agent?*	1.33	1.73	2.31
2.13 To what extent did you feel you could interact with the agent?	-1.11	1.54	2.17
2.14 How much control over the interaction with the agent did you feel you had? (Condition B)**	-1.22	NA	3.35

Table 2: Questions based on the Temple Present Inventory (*p<0.05; **p<0.01).

[1] The complete statistics table can be found here: http://bit.ly/illusion-of-touch-user-study

References

1. Beyer, G., Alt, F., Müller, J., Schmidt, A., Isakovic, K., Klose, S., Schiewe, M., and Haulsen, I. Audience behavior around large interactive cylindrical screens. In Proc. of CHI'11, ACM (New York, USA, 2011), 1021-1030.
2. Brignull, H. and Rogers, Y. 2003. Enticing people to interact with large public displays in public spaces. In INTERACT'03.
3. Cassell, J., Sullivan, H., Prevost, S., and Churchill, E. 2000. Embodied Conversational Agents. Masschusetts, USA : MIT Press.
4. Christian, A. D. and Avery, B. L. 2000. Speak out and annoy someone: Experiences with intelligent kiosks. In Proceedings of the CHI 2000, 313-320, 2000.
5. Glas, N. and Pelachaud, C. "Definitions of Engagement in Human-Agent Interaction". *Workshop on Engagement in Human Computer Interaction (ENHANCE - ACII)*. 2015.
6. Lombard, M., Weinstein, L., and Ditton, T. Measuring telepresence: The validity of the temple presence inventory (TPI) in a gaming context. In ISPR, 2011.
7. Marsella, S. and Gratch, J. "EMA A process model of appraisal dynamics", Journal of Cognitive Systems Research, vol. 10, no. 1, 2009, pp. 70-90.
8. Müller, J., Walter, R., Bailly, G., Nischt, M. and Alt, F. 2012. Looking glass: A field study on noticing interactivity of a shop window. In Proc. of CHI'12, ACM Press, 2012.
9. Peters C., Pelachaud, C., Bevacqua, E., and Mancini, M. 2005. Engagement capabilities for ECAs. 2005. In Fourth International Joint Conference on Autonomous Agents & Multi-Agent Systems, AAMAS '05, Workshop: Creating bonds with humanoids, 2005.
10. Poggi, I. Mind, hands, face and body: a goal and belief view of multimodal communication. Weidler, 2007.
11. Reeves, B. and Nass, C. The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press, New York, USA, 1996.
12. Reitberger, W., Meschtscherjakov, A., Mirlacher, T., Scherndl, T., Huber, H., and Tscheligi, M. 2009. A Persuasive Interactive Mannequin for Shop Windows. In Proceedings of the 4th International Conference on Persuasive Technology, pp. 1-8.
13. Sidner, C.L., Kidd, C.D., Lee, C.H., Lesh, N.B. Where to Look: A Study of Human-Robot Engagement. ACM International Conference on Intelligent User Interfaces (IUI), ISBN: 1-58113-815-6, pp. 78-84, January 2004.
14. Tomitsch, M., Ackad, C., Dawson, O., Hespanhol, L. and Kay, J. 2014. Who cares about the content? An analysis of playful behaviour at a public display. In Proc. PerDis '14. ACM, New York, NY, USA, 160-165.
15. Van Erp, J. and Toet, A. 2013. How to touch humans: Guidelines for social agents and robots that can touch. In Humaine Association Conference on Affective Computing and Intelligent Interaction.
16. Walter, R., Bailly, G., and Müller, J. 2013. StrikeAPose: Revealing mid-air gestures on public displays. In ACM CHI'13 (2013), 841-850.