Shadow Agent: a New Type of Virtual Agent.

Philippe Pasquier
School of Interactive Arts and Technology,
Simon Fraser University
+1-778-782-8546
Vancouver, Canada
pasquier@sfu.ca

Eunjung Han
University of British Columbia
+1-778-866-4477,
8-160 Pembina st. New Westminster, Canada
hanej@ssu.ac.kr

Kirak Kim
HCI Lab., School of Media,
College of Information Science
Soongsil University
+82-2-812-7520,
Seoul, South Korea
raks@ssu.ac.kr

Keechul Jung
HCI Lab., School of Media,
College of Information Science
Soongsil University
+82-2-812-7520,
Seoul South Korea
kcjung@ssu.ac.kr

ABSTRACT
The shadow agent is a new type of intelligent virtual agent exploiting the metaphor of the shadow as a medium. The core of the shadow agent is a behavioral architecture inspired by the BDI (Belief, Desire and Intention) cognitive agent model. The shadow agent perceives the world through two video cameras and a microphone. Advanced computer graphics techniques are used to locate the user's feet and analyses his/her behavior. The shadow agent is embodied as an animated silhouette projected on the floor using the Everywhere Display system and is endowed with simple sonic behavior. While presented as an interactive installation, this model is thought to be a generic approach that can support many more applications, including: video games, museum guides and digital performance to name a few.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – intelligent agents.

General Terms
Algorithms, Performance, Design

Keywords
Virtual Agent, Embodied Agents, BDI model

1.Introduction and Motivations
Artificial agents have been defined as computer systems capable of flexible autonomous action in some environment in order to meet their design objectives [4]. That definition encompasses the following properties:

- *Situatedness* refers to the agent's abilities to perceive the environment via sensors and to act via effectors;
- *Autonomy* refers to the agent's control over its internal state and its capability of action without external intervention;
- *Flexibility* can be broken down in three dimensions:
  - *Responsiveness* refers to the agent ability to respond in a timely fashion to changes in its environment;
  - *Pro-activity* refers to the agent's capability to accomplish goal oriented actions which go beyond simple response to stimulus;
  - *Sociability* refers to the agent's ability to interact with other agents and/or humans.

While the idea of agents takes roots in cognitive sciences it also corresponds to the evolution of computer science programming paradigms. In particular, agent-oriented programming extends object oriented programming in an elegant fashion [8]. *Intelligent virtual agents* are a particular type of artificial agents embodied with a graphical front-end or a physical robotic body. Embodied conversational agents are embodied agents capable of engaging in conversations with humans employing the same verbal or non verbal means that humans do. These have been proven useful as a way to progress toward more natural human-computer interactions [1, 13].

The field of virtual agents is animated by a tension between software virtual agents embodied in a graphical representation which movements are limited to the screen's surface and mobile robots which are still very expensive and limited in their expressiveness. With the first solution, the technology is ready, quite advanced, reliable and affordable (duplication of software entities is virtually free), but the embodiment is virtual which brings all sorts of limitations [13] (in particular spatial). In the second case, despite great achievements, autonomous mobile robots are raising many challenges. The current solutions are onerous and limited in expressiveness and robustness. In this paper, we explore a middle ground and present a new approach that aim to bridge the gap between software and hardware virtual agents and sits in-between graphical interfaces and haptic or tangible interfaces.

The shadow agent is a new type of intelligent virtual agent exploiting the metaphor of the shadow as a medium. This work builds on, generalize and refine previous work with artificial shadow projection [3]. The shadow agent consists of an artificial agent modeled after the the beliefs desire and intention (BDI) architecture [4] and embodied in an animated 2D silhouette that is projected on the floor (as a natural shadow would). The shadow agent interacts with the human user by using pre-defined context-sensitive plans. Since verbal communication is an extremely challenging domain, our shadow agent does rely exclusively on non-verbal communication.

Various behavioral models of artificial agents have been proposed and the present work is based on the beliefs desire and intention (BDI) agent architecture [4]. The BDI framework is now a widely used and accepted model for agents. The model is particularly useful for agents that are resource-bounded, and situated in their environments, such as embodied agents. In the BDI framework, intentions are the key component which specify a course of action to achieve a particular goal in a given context. These are commonly reified as plans which are sequences of actions that provide an abstract representation of an agent's behaviors.

While the system is quite generic and a variety of applications can be envisioned, the rest of the paper will present it as a digital interactive art installation. Interactive art installations are works of art in which the audience direct participation is expected if not required. The digital entertainment industry is a growing market that is known to be shaped by new technologies. It is thus an appropriate domain to test new models and prototypes as the one presented below.

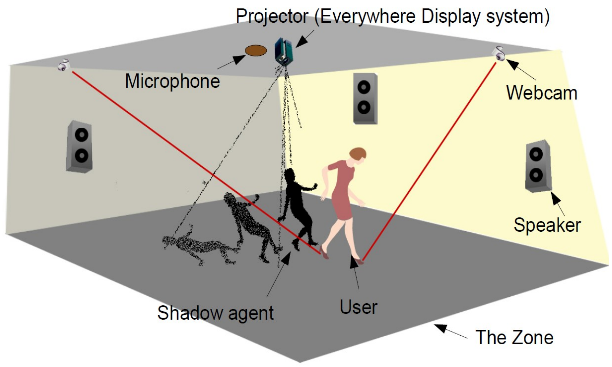


Figure 1. System's hardware components overview.

2.The Shadow Agent Model

The main aim underlying this work is to progress toward more natural human-computer interactions. In particular, we address the seemingly impossible task of developing a willful digital shadow: an otherwise inconspicuous shadow that can, all of a sudden, detach itself from the user and begin acting autonomously, animated by its own (artificial) intelligence. While this idea is simple, the technical realization is not straight forward.

Figure 1 depicts the shadow agent environment. It consists of the following components:

- –The *Zone* is the physical space in which the user and the shadow are located and move;
- –Two *video cameras* (webcams) are used to track the user's feet location as described in previous work [15];
- –The *Everywhere Display* system (itself consisting of a video projector, a motorized mirror and computer graphics algorithms) is used to project the artificial shadow on the floor of the Zone;
- –A *microphone* is used to gather the user's sonic reaction (if any);
- –*Loudspeakers* are used to diffuse sounds emanating from the shadow agent.

Figure 2 presents the overall software architecture of a shadow agent. While the system is not limited in principle, the current system considers a single shadow agent interacting with a single user.

Following the general idea of artificial agents, each shadow agent is characterized by its perception capabilities, a behavioral model and some action capabilities. At the perceptual level, the shadow perceives the world through two webcams and a microphone. However, the shadow's perceptions are treated so that the user's position (and in particular the user's feet position) are known [15]. At the behavioral level, the shadow implements the BDI agent model. Overall, the BDI agent model offers a general cognitive and behavioral model for artificial agents inspired by folk psychology which as proven useful in the development of interactive characters [13]. As such, the shadow has: beliefs (about the user's position, the user's ongoing action, the user's sonic behavior as well as about its own position, ...), desires (instantiated as goals) and intentions (instantiated as plans). Section 3 describes the BDI behavioral model implanted for the shadow agent. At the effectors level, the shadow agent can do a number of actions (moving in the Zone, producing some sound, ...). These actions selected and triggered by the agent's cognition through the BDI model are grouped as context-sensitive plans used to achieve goals. Concretely, the Everywhere Display system is used for all the visual outputs, while 8 speakers are used for the spatialized sonic behavior (described in Section 4).

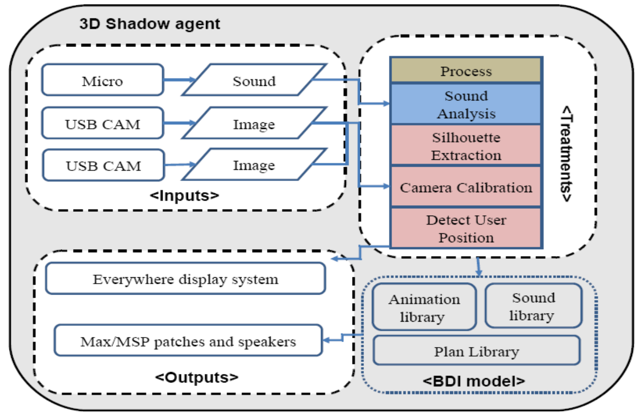


Figure 2. Shadow agent system architecture.

3. The Shadow Agent Behavior

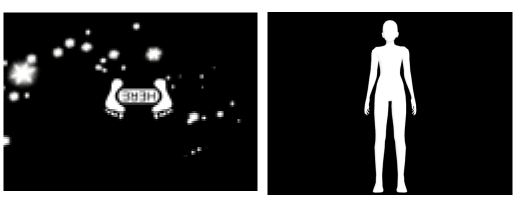
The actual implementation of the BDI agent is using the JACK agent programming framework [9]. In JACK, goals (or desire) are reified as events that are handled by plans (intentions). These can be *internal events* generated by the beliefs base or as part of a plan (instantiating the notion of subgoals) or *external events* corresponding to perceptions (thus allowing for reactive behavior). The bulk of the work when developing an agent is to specify the various plans available along with their context and the triggering events. This section presents the various plans used in the current shadow agent.

The present research purports to encourage the audience to engage in interaction with the shadow agent. Once the surprise triggered by the very presence of the digital shadow is shading, the shadow agent's behavior is critical in achieving this goal. More precisely, depending on the situation, described by the shadow's beliefs base, a number of plans (reifying various intentions) are available (Figures 3 and 4 illustrate some of them):

- –*The waiting plan (Figure 3.a)*: when no user is present in the Zone, then the shadow is absent too and only an image of feet is projected (indicating the starting position).
- –*The apparition plan (Figure 4.a)*: triggered when the user stands on the projected feet of the waiting plan, this plan consists in the shadow appearing and the image of projected feet disappearing. The shadow is growing from the user's feet in an organic "waking up". Birth-like sounds are diffused to attract the user attention. The goal associated with that plan is that the user becomes aware of the shadow's presence.
- –*The immobility plan*: when the user is not moving (probably looking at the shadow) the shadow wiggles its body and observes the user (too).
- –*The following plan*: when the user is moving, the shadow agent will follow him (as a natural shadow would). This plan is active the first minute during which the user discovers the

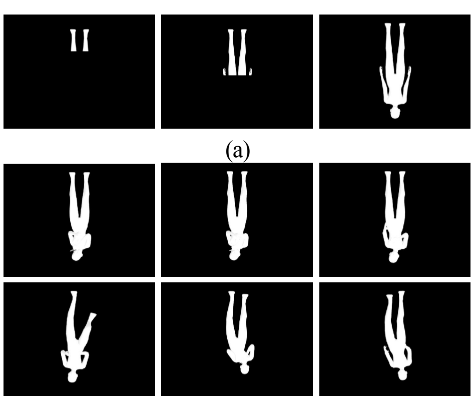
shadow's presence. This conditions the user and prepares him for the next surprise.

– *The moving away plan*: this plan is triggered by the absence of user movement for a fixed amount of time (12 seconds in the current implementation). The shadow starts to move away from the user. The shadow uses a variety of gestures and sounds to focus the user’s attention on itself. In turn, the user is expected to react to the shadow’s activity and becomes more interested in the piece.
– *The call plan*: this plan is triggered when the shadow is too far from the user (4 meters in the current implementation), the shadow will try to call the user and ask her to come by. The shadow body language is used for that.
– *The answering and teasing plan*: if ever the user starts talking to the shadow (this is not unusual), then the shadow will answer to her by generating some teasing noises and movements.
– *The escape plan*: this plan is triggered if the user tries to step over the shadow. It consists in moving away from the user in a swift manner. A sound of discomfort is also played.
– *The dancing plan*: this ludic and self-explanatory plan is triggered when the user is moving in what looks like a pattern (e.g., going back and forth to test the shadow's reaction). A groovy underscore music is played in the background.
– *The sleeping plan*: when the user’s interest begins to wean from the shadow, it acts tired and starts sleeping. If the user approaches and “touch” the shadow at this point, the shadow wakes up again while if the user gets out of the Zone the system is re-initialized.
– *The reinitialization plan*: triggered whenever the user is getting out of the Zone the Shadow looks disappointed and goes back to the middle of the Zone falls asleep and disappear which triggers the waiting plan.



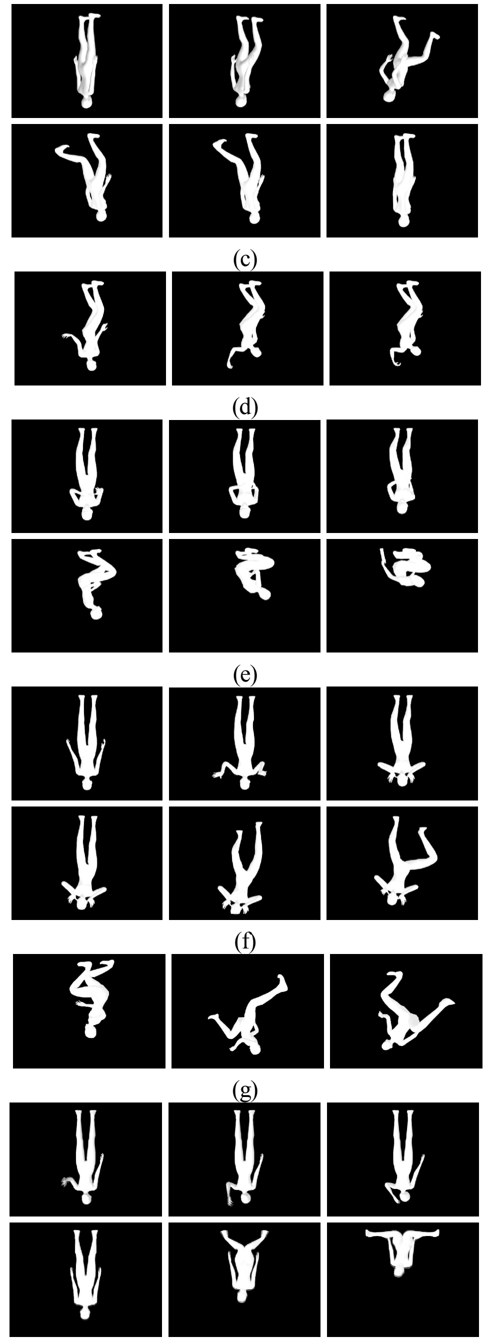
(a) (b)

Figure. 3. (a) Animation of the waiting plan, (b) the basic 2D silhouette used for the animations.



(a)

(b)



(c)

(d)

(e)

(f)

(g)

(h)

Figure. 4. Stills exemplifying some of the animations used for the shadow agent: (a) apparition plan, (b) feet fixed, the immobility plan, (c) the following plan, (d) the call plan, (e) the teasing plan (the shadow laughs, holding its stomach or timidly pokes at the user using a stick), (f) variations of the teasing plan (waving its hands and twisting its body like a snake), (g) the escape plan, (h) The sleeping plan (the shadow yawns, then falls asleep on the floor)

4.Sonic behavior

The shadow agent expresses itself not only visually but also in the sonic realm. Informed by the techniques of sound design for moving images, we decided to give the shadow an abstract sonic vocabulary. Instead of using proper oral language (through the use of text-to-speech system or pre-recorded voice samples), the shadow is mainly using expressive mouth sounds like: breathing, gasping, yawning, laughing, screaming, etc. These sounds achieve both a descriptive function by which they emphasize and highlight the shadow's actions and an interactive function. They are therefore produced in two types of circumstances (as illustrated in some the plans presented in the previous section): (1) they can be attached to the relevant plan in order to emphasize the emotions or (2) they can be used to react to users noise or actions to emphasize the responsiveness of the shadow.

As for the interactive function, the shadow agent perceives the sound through a microphone that is capturing the sonic activity in the Zone. While using voice recognition (a technology still hardly reliable when used without user-specific calibration) is certainly in our future work ambitions, in its current implementation, the shadow reacts to the sound made by the user without interpreting it at the semantic level. The sound perceived includes all the voice sounds (exclamations, comments, laughs, or other noises if any) as well as the ambient soundscape. The main idea is to increase the sensation of presence of the shadow by allowing it to "answer" or "comment" the user's sonic actions, in the aural dimension. This gives the user the sensation to be listened to (which is indeed the case). It was observed that some users literally start talking to the shadow (having in mind that the shadow understands them). This is certainly a phenomenon to be exploited in future work

Concretely, the sound component has been developed using the MAX/MSP signal processing environment [12] and is communicating with the behavioral component using OSC (Open Sound Control) [11]. The sound spatialization is done through volume control over eight speakers (a 7.1 system) and aims to enhance the precision of the shadow's localization.

5. Conclusion

In the present paper, we have presented and discussed the shadow agent model and prototype, a new type of embodied agent architecture. This work is advancing the state of the art in the fields of virtual agents [5], exertion interfaces [10], ubiquitous interactive graphics [7], augmented reality [2], as well as in the broader field of human computer interfaces [1, 13]. In particular, the shadow agent brings the gap between screen-based virtual agent and physically embodied agents (robots) by materializing the shadow agent on the floor (outside of the usual screen surfaces without being a tangible or haptic interface either). The Everywhere Display system allows us to move the shadow agent to any location without loosing definition.

We use advanced computer vision techniques to localize and recognize the user's actions. In tandem with simple audio listening, these inputs are exploited by a BDI agent that acts autonomously by implementing various context sensitive plans in reaction to the user's actions (or the absence of). Camera calibration techniques are used to accurately align the shadow to the user's foot. The visual output is augmented with an audio dimension aiming at accentuating the shadow agent's presence.

The only previous related work know to the authors is an art installation by Zack Booth Simpson and Adam Frank using the idea of shadow projection [6]. Unlike this previous experimental piece, we presented a generic model of shadow agent that is not limited to interactive arts, but can also be applied in many other situations. Examples include: interactive arts and video games, digital performance, museum guides as well as multimedia presentations. We do believe that the synergy between advanced computer graphics and agent technologies is a promising avenue of research as demonstrated by the number of future works for which the presented model lays the ground.

6. References

[1] Jenny, P. 1994. Human Computer Interaction. Addison-Wesley.

[2] Azuma, T. R. 1997. A Survey of Augmented Reality. In *Presence: Tele-operators and Virtual Environments*, Vol.6, No.4, 355-385.

[3] Kim, K., Kwon, W., Choi, E., Park, J., Jeon, M., Ju, H., Choi, H., Baek, W., Kyoung, D., and Jung, K. 2007. The dream of Peter Pan: user interaction arts using computer vision techniques. In *Proceedings of the international Conference on Advances in Computer Entertainment Technology.* ACE '07, vol. 203. ACM, New York, NY, 260-261.

[4] Rao, A. S. and Georgeff, M. P. 1991. Modeling rational agents within a BDI-architecture. In *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, 473–484. Morgan Kaufmann Publishers: San Mateo, CA.

[5] Wooldridge, M. 2001. An Introduction to MultiAgent Systems. Wiley.

[6] Booth Simpson, Z., Frank, A. 2003. Shadow, Eyebeam, New York, Exhibition.

[7] Pinhanez, C. et al. 2002. Ubiquitous Interactive Graphics, IBM research report RC22495, TJ Watson Research Center, New York, USA.

[8] Bordini R., Dastani, M., Dix J. and El Fallah Segrouchni A., Editors, 2007. Programming Multi-Agent Systems (ProMAS), Lecture Notes in Artificial Intelligence, LNAI, vol. 4411, Springer Verlag.

[9] Busetta, P., Ronnquist. R., Hodgson, A., Lucas, A. 1999. Jack intelligent agents - components for intelligent agents in java. AgentLink, News Letter, January 1999. White paper,

[10] Mueller, F., Agamanolis, S., and Picard, R. 2003. Exertion interfaces: sports over a distance for social bonding and fun. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, 561-568

[11] Wright, M., Freed, A., and Momeni, A. 2003. OpenSound Control: state of the art. In *Proceedings of the Conference on New interfaces For Musical Expression.* 153-160.

[12] Puckette, M. 2002. Max at Seventeen, Computer Music Journal, Vol. 26, No. 4, Pages 31-43.

[13] Sharp, H., Rogers, Y. and Preece, J. 2007. Interaction Design: Beyond Human–Computer Interaction, 2nd ed. John Wiley & Sons Ltd.

[14] Norling, E. and Sonenberg, L. 2004. Creating Interactive Characters with BDI agents. In *Australian Workshop on Interactive Entertainment,* Sydney, Australia.

[15] Donwuk, K. Yunli, L. Woonhyuk, B. Eunjung, H. and Keechul, J. 2006. Efficient 3D Voxel Reconstruction using Precomputing method for Gesture Recognition, Proceeding of *First Korea-Japan Joint Workshop on Pattern Recognition*, pp. 67-72.