Towards Engaging Full-Body Interaction

Paolo Petta[1], Alexander Staller[1], Robert Trappl[1,2],
Stephan Mantler[3], Zsolt Szalavári[3], Thomas Psik[3], Michael Gervautz[3]
[1]Austrian Research Institute for Artificial Intelligence
[2]Dept. Medical Cybernetics and Artificial Intelligence, University of Vienna
[3]Institute of Computer Graphics, Vienna University of Technology

Scenario

We implemented an interactive virtual environment based on the "magic mirror" metaphor introduced with the MIT ALIVE project MIT (Darrell et al. 1994). While this technology has been used in other work to research agent architectures based on ethological models (Blumberg 1997) or scripted computer theatre (Pinhanez and Bobick 1998), our permanent exhibit in the Vienna Museum of Technology aims at taking advantage of the possibility opened up by this particular kind of unencumbered immersion in a virtual environment for the users to bring in their rich expertise in full-body action and communication, so as to provide a satisfying and truly interactive experience for laypersons. To this end, the system comprises a synthetic character, the *Invisible Person*, designed to improvise with the visitors.

Evidence from the deployment of public virtual environments under comparable circumstances indicates that the use of highly specialized domains faces limited success (cf. e.g. (Martinho and Paiva, 1999)). This is both because of the high expertise required to discern the often subtle changes, as well as due to the overly long amount of time that has to be invested in order to get accustomed with the system. In order to avoid these problems and also to mitigate the effect of the inevitable limitations of systems dependent on purely vision-based user tracking (e.g. the unreliable recognition of postures or gestures), the Invisible Person was designed to be suggestive of adequate kinds of interaction. Its lack of a face or fingered hands relates to the coarse resolution of the vision, while its child-size appearance shall encourage sequences of short and well-defined actions (as well as lower expectations of perfect recognition). Together with the employment of high-quality motion-captured animation, this balance between the full capabilities of the agents and the capabilities elicited from the human user contributes to make the exhibit interesting to visit. In addition, the simple scenario—a bare stage and a single synthetic actor—also opened up the opportunity to deploy and evaluate a situated implementation of the functional appraisal theory of emotions to contribute to the solution of the action expression problem (Petta 1999).

A principled Emotion Architecture

An essential ingredient for the exhibit's success is the endowment of the agent with interaction skills, clearly identifiable personality traits, and emotional competencies. Our line of research builds substantially upon research on the appraisal theory of emotions (Frijda 1986, Smith et al. 1996). Herein, the emotional is modeled as a dual control process for the satisfaction of an individual's *concerns*. Appraisal of events leads to the generation of *action tendencies*, the readiness to achieve or maintain a given kind of relationship with the environment. Emotions are then defined as *changes* in modes of relational action readiness, either in the form of tendencies to establish, maintain, or disrupt a relationship with the environment, or in the unbounded mode of relational readiness as such. This approach thus directly connects characteristic expressive behaviour to its underlying principles, the related governing action tendencies. As emotions in turn determine the organization and patterning of the individual's functioning, they effectively form its personality, determining its identity. This identity finally governs the individual's way of interacting with the environment. Thus, a limited set of action tendencies and concerns ought to be both necessary and sufficient for the principled realization of a synthetic character's personality.

The Invisible Person is based on a three-layer architecture (Bonasso et al. 1997) with the addition of a tightly integrated emotional subsystem that allows appraisals to occur at all processing levels (Reekum and Scherer, 1997, Staller and Petta 1998). Generation and management of the final action tendencies is located at the middle layer, from where the scheduler controls the system in a "middle-out" fashion (cf. Gat 1997) triggering activation of reactive plans in the top level, and selecting and supervising the execution of reactive action packages in the bottom layer. Actions are selected out of a static library of around 150 motion-captured and post-processed basic animation actions (such as "one step forward" or "knee down", but also "scratch your head" or even "take a picture of the visitor") which may be scaled in speed, rotation, and translation. The intended action is posted to the animation subsystem which manages the animation pipeline and posts back the current state (location and orientation) of the Invisible Person. As the net result of a given sequence of basic animations depends on a number of parameters (such as current system frame rate or the scaling parameters of the single animations), the system shares many characteristics of "real" robotic systems. This is the second main reason for the use of a trionic architecture. The second main channel of action expression—besides the dynamics of behaviour execution—is provided by full-body animated textures that encode the current level of different concerns and the degree of activation of action tendencies as transmitted independently from the stream of basic actions from the trionic agent architecture to the animation subsystem. The encoding utilizes perceptive dimensions such as colour tone, "excitedness" of the pattern, or grain and (rotational and translational) speed of the texture movement across the body. The texture may alter rather swiftly upon changes of action tendency, often anticipating actual physical action: this is but one example for "life-like" effects that were not explicitly designed for but account for some of the success the exhibit has found immediately.

The Exhibit

The system, that runs on two Silicon Graphics computers (an O2 for the vision subsystem and an Octane hosting agent architecture and computer animation) has been a major attraction of the museum since its re-opening in June, 1999. Currently, a first evaluation is underway to determine directions of future improvements. Likely candidates include better coverage of specific situations (multiple visitors, children) and implementation of short- and long-term adaptive strategies. Especially for the latter, the generative approach of the employed emotion model is expected to go a long way towards ensuring preservation of consistent behaviour even with changing system parameters, obviating the need of complex re-engineering as would be required for "reified" shallow emotion models.

Acknowledgement: The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry of Science and Transport.

References

Blumberg, B. M. (1997). Old Tricks, New Dogs: Ethology and Interactive Creatures. Ph.D.Thesis, MIT, Cambridge, MA, USA.

Bonasso, R. P., Firby, R. J., Gat, E., Kortenkamp, D., Miller, D. P., Slack, M. G. (1997). Experiences with an Architecture for Intelligent Reactive Agents. In Hexmoor H. (Ed.), Special Issue: *Software Architectures for Hardware Agents*, Journal of Experimental and Theoretical Artificial Intelligence, 9(2/3), pp. 237-256.

Darrell, T., Maes, P., Blumberg, B. & Pentland A. P. (1994). A Novel Environment for Situated Vision and Behavior. MIT Media Laboratory, Perceptual Computing Section Technical Report No.261.

Frijda, N. H. (1986). *The Emotions*. Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris.

Gat, E. (1997). On Three-Layer Architectures. In Kortenkamp, D., Bonasso, R.P.& Murphy R. (Eds.): *Artificial Intelligence and Mobile Robots*, MIT/AAAI Press.

Martinho, C. & Paiva, A. (1999). Developing Pathematic Agents. In Velasquez, J. D. (Ed.): "Emotion-Based Agent Architectures" (EBAA'99, Seattle, USA, May 1, 1999), pp.79-83.

Petta, P. (1999): Principled Generation of Expressive Behavior in an Interactive Exhibit. In Velasquez, J. D. (Ed.): "Emotion-Based Agent Architectures" (EBAA'99, Seattle, USA, May 1, 1999), pp. 94-98.

Pinhanez, C. S. & Bobick, A. (1998). "It/I'": A Theater Play Featuring an Autonomous Computer Graphics Character, MIT Media Laboratory, Perceptual Computing Section Technical Report No.455.

Reekum, C. M. van & Scherer K. R. (1997). Levels of Processing in Emotion-Antecedent Appraisal. In Matthews, G. (Ed.), *Cognitive Science Perspectives on Personality and Emotion*, Elsevier, Amsterdam, pp.259-300.

Smith, C. A., Griner, L. A., Kirby, L. D., Scott, H. S. (1996). Toward a Process Model of Appraisal in Emotion. In Frijda N. H. (Ed.): *Proc. of 9th Conf. of the International Society for Research on Emotions* (ISRE'96, Toronto), pp. 191-105.

Staller, A. & Petta P. (1998). Towards a Tractable Appraisal-Based Architecture for Situated Cognizers. In Canamero, D., Numaoka, C. & Petta, P. (Eds.), *Grounding Emotions in Adaptive Systems*, Workshop Notes, Fifth International Conference of the Society for Adaptive Behaviour (SAB98, August 21, 1998) Zurich, Switzerland.