Principled generation of expressive behavior in an interactive exhibit

Paolo Petta
Austrian Research Institute for Artificial Intelligence (ÖFAI)
Schottengasse 3, A-1010 Vienna, Austria
paolo@ai.univie.ac.at

1 Introduction

The present work was carried out in the context of the development of an immersive interactive virtual environment in which a single human user and a synthetic actor ("The Invisible Person") engage in an improvisational interaction between equally entitled peers (Figure 1). The limitations of this scenario with respect to the number of parties involved and the synthetic actor's perceptual and communicative capabilities allow to experiment with the modeling of expressive behavior, a noted deficiency of early synthetic actor agents that has been recently topicalized as the "action expression" problem ([15]).

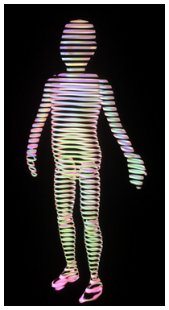

Figure 1: The Invisible Person

The paper is structured as follows: the next section describes the interactive scenario in more detail. We next characterize the action expression problem and then proceed to present our solution approach. We relate results from research in software architectures for embodied agents and the appraisal theory theory of emotions and investigate their suitability as a principled basis for behavior expression generation. Section 5 describes the implementation in the agent architecture employed in the present system, and we wrap up our presentation relating the results to our ongoing work in the framework of the TABASCO architecture for emotions [17].

2 Project scenario: an interactive exhibit

In a joint project carried out by the Austrian Research Institute for Artificial Intelligence and the Dept. of Computer Graphics of the Vienna Univ. of Technology, we are working on the implementation of an interactive virtual environment based on the "magic mirror" metaphor introduced with the ALIVE project at MIT [4]: a human user stands in front of a large screen onto which her own image is composited into a display which includes synthesized animated graphics.

The main (logistic) advantage of this approach is that it allows for an untethered device-free, unencumbered immersion of human users. This makes it an attractive choice for a largely unattended permanent exhibit. On the downside there are most prominently limitations deriving from the purely vision-based user tracking and interacting (performed by a single camera): while the location of the user can be assessed rather exactly, recognition of posture or gestures is difficult to perform reliably and topic of intensive research (e.g. [5]). Another downside is given by the fact that the user has to maintain eye-contact with the mirror at all times. In addition to these constraints, experiences with public deployments of virtual environments (e.g. an exhibit at the world EXPO'98 in Lisbon, Portugal, featuring virtual dolphins[12]) show that use of specialized domains finds limited success, both because of the high expertise required to discern the often subtle changes taking place as well as because of the overly long timespan that has to be invested in order to get accustomed with the system.

The present project attempts to sidestep most of these problems by keeping the scenario simple and placing the focus on the very possibility for "laypersons" to bring in their own rich expertise in full-body action and communication so as to realize a satisfying and truly interactive personal experience. The investigated scenario thus comprises a single human user in front of the "magic mirror" and a synthetic actor, the "Invisible Person", who engage in an improvisational interaction between equally entitled peers. The appearance of

the Invisible Person was designed to meet the criteria of immediately suggesting adequate kinds of interaction: the lack of a face and fingered hands is related to the coarse resolution of the vision system and the limitation of the perceptual system, while the child-sized appearance shall encourage sequences of short and well-defined actions (as well as lower expectations of perfect recognition — see e.g. [7] for a pertinent discussion). In this sense, the very simplicity of the scenario is aimed at supporting the maintenance of a balance between the full capabilities of the synthetic actor on one side and the capabilities actually employed by the human user. The expectation is that the resulting emphasis on the subjective experience of "virtual presence", along with the employment of high-quality motion-captured animation, will make the exhibit interesting to visit.

As one caveat we note that while the basic repertory of behaviors of the Invisible Person is composed of actions suitable for broad-and-shallow interaction (such as evading, pursuing, staying close to, circling around) we are currently evaluating the necessity of having to include more precisely defined "scenarios" after all. These should serve as associative facilitators for users lacking spontaneity. For example, having the Invisible Person don a work-out outfit and start "working out", or having it place a virtual Chinese lantern close to the screen and start to cast virtual (full-body) shadows could turn out to be necessary measures to "de-block" certain kinds of users.

For the purpose of the present paper, the very simplicity of the scenario has two implications. On one hand the limited complexity of the single behaviors (required so as to encourage free association and spontaneous improvisation) makes tackling the action expression problem, discussed below, an important issue. On the other, this reduced complexity also motivated us to try to devise a fully algorithmic solution.

3 The action expression problem

In [15], Phoebe Sengers introduces the *Expressivator*, an extension to the Hap architecture [11, 14] designed to tackle the "action expression problem", that occurs in particular in behavior-based agents: even if "dithering" between narrowly competing behaviors is eliminated, they still do tend to abruptly jump from one behavior to another according to their internal action selection principles, which can result in action sequences that are confusing in the spectator's eyes. The action-expression concept is well-known from the character animation area, where it is considered a solved topic. In the animated believable agents domain however, there currently is a lack of principled ways to tackle the problem: what kinds of modifications and extensions are necessary to which part of the architecture, and for what reason?

Sengers recasts the problem as one of management of *external signs*, overt behaviors which communicate the "reasons" for the actions that are carried out by the agent. A sign management system keeps track of what has been communicated to the user and influences action decisions on the basis of the user's probable perception of the agent's state. She highlights the ensuing relevance of *transition behaviors*, which are to convey the rational for behavioral changes to the user. These transition behaviors are implemented in the behavior-based framework provided by Hap as *meta-level controls*: "In addition to supporting transitions, meta-level controls can express to the user parts of the agent architecture that were formerly implicit (and therefore invisible)" [15], p.25.

The Hap architecture is based on a stripped-down version of the PRS system [11]. Basically, it comprises a static library of plan trees and a rather straightforward execution algorithm which selects the most specific currently enabled node of a tree for expansion (if an internal node) or execution (if a leaf). We mention this detail, because the circumstance that this approach proved sufficient at the time of the implementation of Tok (the reactive planning component of Hap) at that time obviated the need to include a *scheduler*, by now a standard component of the established three-layer architectures for embodied agents [10, 9]. Similar deficiencies are also shared by other notable architectures, such as Hamsterdam [2].

It is then not surprising that Sengers mentions the difficulties encountered also by other researches in expressing the relationships *between* individual behaviors, given their essential encapsulation and modularity: "behaviors do not know enough about other behaviors to be able to express to the user their interrelationships" ([15] p.27). A difficult resulting issue is what precisely this "enough" should refer to. The solution implemented in the Expressivator is the provision of explicit connections between behaviors represented by a special class of *transition behaviors*. These are to explain why the agent's behavior is changing and what its intentions are: "Instead of simply engaging in apparent stimulus-response activity, the agent shows that there are reasons for its behavioral decisions, thereby giving the user more insight into its motivations. Transition behaviors show that the agent is truly conscious, a thinking being that considers, however briefly, before it acts." ([15] p.27). Since users are assumed to understand behaviors at a higher level, only such transitions are deemed necessary that express what the agent is "fundamentally" doing, e.g. eat, play or sleep[1].

In her "initial foray into the land of action-expression", Sengers presents an empirical list of 12 transition types and 7 meta-level controls used to implement them on top of Hap. She asserts that the "reduction of modularization" resulting from the connection of behaviors in the offered solution is a notable price to pay for an extension that is seen to be aimed at the *quality* rather than the *correctness* of the behavior.

[1] Rapid change between these high-level behaviors is expected to be prevented by the "dithering-control" mentioned before.

4 Towards a principled solution

In this section we relate the state of affairs reported in the previous one to results from appraisal theory of emotions and layered agent control architectures, and suggest that these domains provide adequate means for attacking the action expression problem in a principled way.

4.1 Contributions of appraisal theories

"Appraisal theories" of emotions postulate the existence of processes that continuously evaluate the whole environment (which comprises the "external" environment and the subject itself) according to dimensions of relevance for the individual. These evaluation processes are called appraisals. They are taken to mediate between the occurrence of significant events and changes of internal "action tendencies" that are accompanied by characteristic expressive behavior. An action tendency is defined as readiness for different actions having the same intent or goal state [8] pp.70-71, i.e. "states of readiness to achieve or maintain a given kind of relationship with the environment". Appraisals thereby provide an explanation why the same event can give rise to different emotions in different individuals, or even in one and the same individual at different times. Conversely, appraisals offer an explanation for understanding what differentiates emotions from each other. Emotions are thusly 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 form of mode of relational readiness as such. [8] p.71[2].

Nico Frijda [8] p.73 distinguishes between two different principles of categorization of emotion: by action tendency change and by nature of the emotional object. The latter categorization inevitably is highly dependent upon which objects are being distinguished and considered important by the environment providing the categorization. One such categorization was published in [13] — subsequently extended by Clark Elliott ([6] and following) — and quickly advanced to become the most popular "reference model" of appraisal used in architectures for synthetic agents, including Hap. Frijda also states an important characteristic of this second of the two approaches: "As consequence of this dual principle of categorization, emotions defined primarily by their object cannot be specified by action tendency or activation mode. This implied that they have no characteristic facial expression and that their presence cannot be recognized by means of expressive behavior alone. ([8] p.73).

It thus should not come as a surprise that "action expression" becomes a particular problem in architectures that are based solely on this second principle. Differently from what applies to the second categorization, there exists a list of action tendencies (and activation modes) that appear to be elementary, not composites. Each of these is conceptually distinct, in terms of a particular relational aim or sense. Each of these appears to correspond with a species-specific behavior mode or system, in humans and other higher animals, or otherwise with explicit non-occurrence of a particular behavior mode ([8] p.87).

To summarize, for the purposes of the present analysis, action tendencies can be seen to provide a separate classification system defining how the behavioral repertory of an agent can be organized in classes that share specific expressive characteristics. Even for novel "virtual" creatures and their environments, these action tendencies can be derived in a principled way from an analysis of their particular lifeworld [1]. To illustrate, the following could be mappings of the behavior sequences used as examples in [15]:

An agent is assumed to be currently napping but then deciding to start exercising. The possible reasons stated are: ([15] p.26)

1. It could be well-rested and ready for something strenuous.
2. It could feel guilty about napping because it was trying to stay in shape
3. It could be engaging in an exercise marathon, but just work up after accidentally falling asleep in the middle of the marathon
4. It could be threatened by another agent, who is forcing it to exercise against its will.

A related action tendencies mapping could be: 1 — free activation, 2 — submitting, 3 — agonistic, 4 — rejecting. (cf. footnote 2). In the presented view, the problem of "action expression" thus naturally folds into the issue of the realization of a "whole" model of the appraisal process, as an aspect of translating *transitions* between action tendencies into overt action. In this sense, action expression itself becomes part of a *correct* modeling of behavior.

4.2 Lessons from control architecture design

Substantial progress has been made in the area of design of software architectures for embodied agents, in particular with respect to layered designs [10, 9]. Out of the many noteworthy results, we propose that the change in appreciation of the role of the "middle" tier of the well established "trionic" three-layer model, the *scheduler*, is of particular relevance for the problem at hand: among other characteristics, we think it provides a natural location for the implementation of at least a substantial part of the functionality of Senger's Expressivator. In this respect we would like to particularly remind of the import of the *cognizant failure* of behaviors in order to insure that the scheduler can dispose of all required information, including the one stemming from the lower layer [9]: given the impossibility of preventing all possible failure causes, this approach takes the alternative route of providing information about the supposed reason why a behavior failed to execute.

[2] We have to refer the interested reader to the cited literature for a precise definition of the terms occurring in this definition.

5 The Invisible Person

For the purposes of the present application, the already discussed broad and shallow nature of the exhibit under development provided incentives in favour of trying to go all the way towards a fully algorithmic solution of action expression. From another perspective, this exploitation of opportunities is allowing us to start gathering first hands-on experiences with respect to the considerations that led us to the theoretical design of the appraisal-based TABASCO architecture for emotions, which more fully relates the layered information processing models being developed in robotics, psychology and neurobiology [17].

5.1 Action generation

The Invisible Person is driven by a control architecture similar to Tok [11], but which includes a distinct scheduler that form the nexus between the low-level execution system and an additional regulatory concern (i.e, goal) satisfaction system[3].

After each invocation of a current behavior, external information comprising data about the user (location, movement, posture, etc.) and discrete events signaled by the geometry system (e.g., collisions) and internally gathered information and computed statistics about the system performance are "appraised" and related to the agent's current concerns (long-term goals such as constraining modeled fatigue as well as current goals = selected behavior trees). The thereby instantiated action tendencies provide additional context both for active behaviors, e.g. indications "how hard to try" to establish or maintain the related preconditions, and the scheduler itself, e.g. the relevance of aspects such as "competence/coping potential" (= arity of a high-level behavior tree candidate) or "self-control" (number of open choices) for high-level behavior selection.

5.2 Action Expression

As described above, the Invisible Person does not have a proper "face"— this design choice resulted among others from the "symmetry" criterion of trying not to model aspects of the agents that cannot in turn be perceived from the human user. As a surrogate, the main means of action expression, apart from the dynamics of behavior execution, is being implemented in the form of animated textures which encode the activation of action tendencies in dimensions of psychological perception such as warmth of the color tone, feel of the material or "excitedness" of the texture pattern. This mode of appearance thus is subject to gradual change during pursuit of single patterns and thereby announces e.g. the approach of activation thresholds of action tendencies. In order to allow for full appreciation of texture-encoded information at all times, care had to be taken to differentiate between aspects that would be of relevance in states of high activity and such that apply to conditions of relative passivity and their respective encoding, given that e.g. pattern changes are hard to discriminate on a moving body.

The "activation" of the action tendencies is modeled after the laws stated in [8]. The exact quantitative parameter assignments for the qualitatively described dynamics is part of the tuning process currently underway.

[3] Additional interesting parallels between the emotion system's capability of "spawning" long-term planning processes theorized in the psychological literature on appraisal on one hand and the distinction of the middle sequencing layer as center of a "middle-out" model of control e.g. in Gat's ATLANTIS architecture will be the covered in another publication. See also [16] as another recent example of a synthesis that adopts the "trionic" layout.

6 Conclusion

We have presented our principled line of attack to the action expression problem for animated synthetic characters: we related empirical findings reported in earlier publications to established results in the domains of cognitive psychology and software control architectures for embodied agents. While preliminary testing is providing positive empirical feedback for our design decisions, the full deployment and analysis of results has to be deferred to a later publication.

We readily admit that terming the modeled architecture "appraisal-based" at the present state of implementing functionality can only be an indication of intended further development, at best: we are planning to incrementally continue to extend the current implementation, with respect of the agent's competence and complexity, inclusion of further aspects of appraisal as information processing process (cf. [17]), but also modeling of physiological characteristics (e.g. [3], [18]), thereby working our way towards the realization of a comprehensive principled and grounded model of the emotional.

7 Acknowledgements

The reported research was carried out in the context of a project funded by the Vienna Museum of Technology. The author would like to thank Robert Trappl, Alexander Staller, Michael Gervautz, Stephan Mantler and Zsolt Szalavari of the project team for their support in the authoring of this paper, and another particular friend for the incomparable motivational work.

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

References

[1] Agre P., Horswill I.: Lifeworld Analysis, Journal of Artificial Intelligence Research, 6:111-145, 1997.

[2] Blumberg B.M.: Old Tricks, New Dogs: Ethology and Interactive Creatures, Massachusetts Institute of Technology, Cambridge, MA, Ph.D.Thesis, 1997.

[3] Canamero D.: Modeling Motivations and Emotions as a Basis for Intelligent Behavior, in Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, CA, USA, February 5-8, ACM Press, pp.148-155, 1997.

[4] Darrell T., Maes P., Blumberg B., Pentland A.P.: A Novel Environment for Situated Vision and Behavior, MIT Media Laboratory, Cambridge, MA, Perceptual Computing TR No.261, 1994.

[5] Eickeler S., Kosmala A., Rigoll G.: Hidden Markov Model Based Continuous Online Gesture Recognition, Proc. 14th Int'l Conf. on Pattern Recognition (ICPR'98), Brisbane, Australia, August 16-20, 1206-1208, 1998.

[6] Elliott C.D.: The Affective Reasoner: A process model of emotions in a multi-agent system, Northwestern University, Illinois, Ph.D.thesis, 1992.

[7] Foner L.N.: Entertaining Agents: A Sociological Case Study, in Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, CA, USA, February 5-8, ACM Press, pp.122-129, 1997.

[8] Frijda N.H.: The Emotions, Cambridge University Press, Editions de la Maison des Sciences de l'Homme, Paris, 1986.

[9] Gat E.: On Three-Layer Architectures, in Kortenkamp D., Bonasso R.P., Murphy R. (eds.): Artificial Intelligence and Mobile Robots, MIT/AAAI Press, 1997.

[10] Hexmoor H.: Special Issue: Software Architectures for Hardware Agents, Journal of Experimental and Theoretical Artificial Intelligence, 9(2/3), 1997.

[11] Loyall A.B.: Believable Agents: Building Interactive Personalities, Carnegie- Mellon University, Pittsburgh, PA, Ph.D.Thesis, 1997.

[12] Martinho C., Paiva A.: Pathematic Agents, Proceedings of Autonomous Agents'99, Seattle, USA, forthcoming.

[13] Ortony A., Clore G.L., Collins A.: The Cognitive Structure of Emotions, Cambridge University Press, Cambridge, UK, 1988.

[14] Reilly W.S.N.: Believable Social and Emotional Agents, School of Computer Science, Carnegie Mellon University, Ph.D.Thesis, TR CMU-CS-96-138, 1996.

[15] Sengers P.: Do the thing right: an architecture for action-expression, in Proceedings of the second international conference on Autonomous agents (Agents'98), Minneapolis/St. Paul, MN, USA, May 9-13, ACM, New York, pp.24-31, 1998.

[16] Sloman A.: Why can't a goldfish long for its mother? Architectural prerequisites for various types of emotions., Conference on Affective Computing: The Role of Emotion In HCI, University College London, London, UK, April 10, Invited Talk, 1999.

[17] Staller A., Petta P.: Towards a Tractable Appraisal-Based Architecture for Situated Cognizers, in Canamero D., et al.(eds.), Grounding Emotions in Adaptive Systems, Workshop Notes, 5th International Conference of the Society for Adaptive Behavior (SAB98), Zurich, Switzerland, August 21, pp.56-61, 1998.

[18] Velasquez J., Maes P.: Cathexis: A Computational Model of Emotions, in Proceedings of the First International Conference on Autonomous Agents, Marina del Rey, CA, USA, February 5-8, ACM Press, pp.518-519, 1997.