We Never Stop Behaving: The Challenge of Specifying and Integrating Continuous Behavior

Hannes Högni Vilhjálmsson[1], Elías Ingi Björgvinsson[1], Hafdís Erla Helgadóttir[1], and Stefán Ólafsson[1]

[1]CADIA, School of Computer Science, Reykjavik University

1 Main Research Themes

The Socially Expressive Computing group aims to build virtual environments where social interaction is both effective and visually convincing. What particularly sets our work apart is the belief that to reach this goal all social bodies, regardless of whether they are avatars representing human users or characters controlled by autonomous agents, need to automate social awarenes and reactivity to the social environment (Vilhjálmsson, 2014).

This paper is an exposition of the idea that it is not trivial to bring typical Embodied Conversational Agent (ECA) architectures from a situation where they support one-on-one conversations with human users into virtual environments populated with other agents and avatars. This has implications for SAIBA because it grew out of the former situation. The following sections briefly review three phases of prior and current research that demonstrate this.

1.1 Phase 1: Discrete Control of Behavior

In the Embodied Conversational Agent REA (Cassell et al., 1999), the architecture was structured around the agent's response to discrete events generated by the actions of a human interlocutor (Figure 1). These multimodal events were fused into so-called frames which contained both interactional and propositional interpretations of what just happened. Those in turn created obligations, addressed by similar outgoing frames which generated schedules of supporting verbal and non-verbal actions.

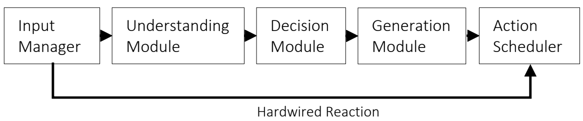

Figure 1: The REA architecture was a pipeline for generating a proper response to a user's input. A short-cut supported a quick hardwired reaction, but in the absence of user input events, nothing would happen

The multi-modal generation process, from text to be spoken to annotated intent to supporting behavior, got consolidated into a flexible tool called BEAT (Cassell et al., 2001). This was an early instance of FML and BML annotation and scheduling. Spark, which animated avatars based on the online chat messages exchanged by their users, used this approach (Vilhjálmsson, 2004). Avatars would idle until someone ”spoke” and a multimodal performance ensued. The bursts of lively conversation were framed by awkward inactivity.

1.2 Phase 2: Coordinated Social Environment

The Tactical Language and Culture Training System (Johnson et al., 2004) was a virtual environment populated with interactive characters playing roles in scenarios for training human users in the use of foreign languages and cultural skills (Figure 2).

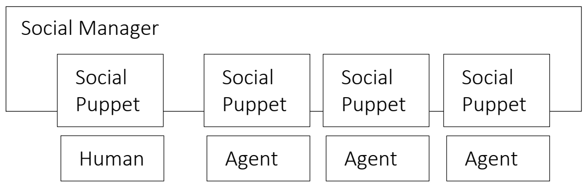

Figure 2: The social puppets of TLCTS could react to any other social puppet in a coordinated fashion thanks to a central manager. Emphasis was on initiating and breaking contact.

Bringing ECAs into an interactive game environment aggrevated the awkwardness inactive moments. People standing idly, waiting to be spoken to, were neither engaging nor natural. A way was needed to naturally manage the co-presence of multiple social bodies that would for example react to someone approaching or even just passing by. This was done with a centralized social manager that kept track of all *social puppets*, which essentially were the embodiments of the agents and users (Vilhjalmsson et al., 2007). The manager fed puppets with perceived FML events, mostly of the interactional kind, such as *A recognizes you* or *A requests turn in your group*. Each puppet would then provide the manager with an FML response, such as *Invite A to join* or *Give A the turn*, as well as animating the supporting nonverbal behavior.

While the social puppets could bounce around FML messages, potentially creating intricate patterns of interactional behavior, each behavior onset or change in idle state (each character could display a variety of idle motions based on context) was a discrete response to a discrete event. Sometimes these meaningful events would occur seconds or even minutes apart. Was no behavior necessary in between?

1.3 Phase 3: Continuous Motion Control and Emergence

People do not simply stop behaving. Our behavior is continuous and is modulated by a constantly changing environment. To capture this notion, it seemed natural to think of the body as being pushed or pulled by a an ever present force, not physical but an abstract one, kind of a *social* force.

Our first instantiation of this, CADIA Populus (Pedica and Vilhjalmsson, 2010) got completely rid of the conversation focused ECAs and neither considered FML or BML. Instead it focused on dynamic positioning and orientation of bodies in large social environments based on a steering behavior framework (Reynolds, 1999). We implemented such things as Kendon's F-formation system (Kendon, 1990) and reactions to the invasion of private space).

At any given moment one or more steering forces would motivate motion in any of the degrees of freedom (forces could be prioritized or combined using weights). These forces would belong to certain contexts, essentially implementing the social norms associated with particular situations. Only those in conversation would for example need to worry about maintaining their F-formation (Figure 3.

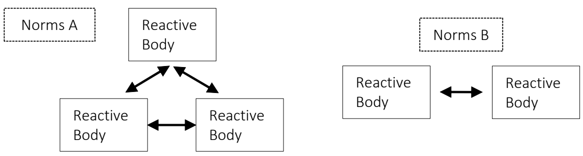

Figure 3: Social bodies are continuously influenced by social forces, tied to certain contexts such as group membership, and update their position and orientation accordingly

This approach resulted in a far more life-like environment (Pedica et al., 2010) where overall social order seemed to emerge from relatively simple rules, independently executed by every character.

2 Current Architectures and Standards

We have now been seeking ways to integrate our continuous steering approach with mechanisms that also support action planning. We have both integrated BDI systems (Thrainsson et al., 2011) and behavior trees (Pedica and Vilhjálmsson, 2012). We are currently combining a REA-like architecture with the behavior tree-based social steering mechanism we call Impulsion. This is taking place in a Unity 3D environment called Virtual Reykjavik (Figure 4). The work is not completed, e.g. our proposed FML (Cafaro et al., to appear) has not been fully integrated with Impulsion.

Figure 4: Groups gather in the central square of Virtual Reykjavik

3 Future Architectures and Standards for IVAs

The challenge that SAIBA faces is that for life-like characters we need a multitude of motion engines (e.g. for locomotion, for gatherings, for "idling" and gazing), that continuously shape nonverbal behavior - *not on a timed schedule*, but as a reaction to a dynamic environment. To some extent we have addressed this by including BML commands that "set" the state of such engines, but we have not specified what then occurs in any detail.

Furthermore, complex motion engines can be valuable assets and it seems that the SAIBA community could work on interfaces that support their migration between behavior realizers. Being able to shop for components such as gaze engines and gathering engines could save a lot of work.

4 Suggestions for Discussion

In light of the previous discussion, the following topics are suggested for the workshop:

- How does SAIBA currently deal with continuous behavior?
- What is our experience with integrating SAIBA with continuous motion engines (e.g. for locomotion)? Is there inherent incompatability?
- Do we see value in addressing continuous motion within the SAIBA framework? Perhaps with a motion engine interface specification?

References

A. Cafaro, H. H. Vilhjalmsson, T. Bickmore, D. Heylen, and C. Pelachaud. to appear. Representing communicative functions in saiba with a unified function markup language. In *Proceedings of the 13th International Conference on Intelligent Virtual Agents*. Springer-Verlag.

J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. H. Vilhjalmsson, and H. Yan. 1999. Embodiment in conversational interfaces: Rea. In *Proceedings of CHI*, pages 520–527. ACM Press.

J. Cassell, H. H. Vilhjálmsson, and T. Bickmore. 2001. Beat: The behavior expression animation toolkit. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 477–486, New York, NY, USA. ACM.

W. L. Johnson, S. Marsella, and H. H. Vilhjalmsson. 2004. The darwars tactical language training system. In *Proceedings of I/ITSEC*. SSA.

A. Kendon. 1990. *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press, Cambridge, England.

C. Pedica and H. H. Vilhjalmsson. 2010. Spontaneous avatar behavior for human territoriality. *Appl. Artif. Intell.*, 24(6):575–593, July.

C. Pedica and H. H. Vilhjálmsson. 2012. Lifelike interactive characters with behavior trees for social territorial intelligence. In *ACM SIGGRAPH 2012 Posters*, SIGGRAPH '12, pages 32:1–32:1, New York, NY, USA. ACM.

C. Pedica, H. H. Vilhjalmsson, and M. Larusdottir. 2010. Avatars in conversation: The importance of simulating territorial behavior. In A. et al., editor, *in Proceedings of the 10th International Conference on Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 336–342. Springer Berlin Heidelberg.

C. W. Reynolds. 1999. Steering behaviors for autonomous characters. In *in the proceedings of Game Developers Conference*, pages 763–782. Miller Freeman Game Group.

P. R. Thrainsson, A. L. Petursson, and H. H. Vilhjálmsson. 2011. Dynamic planning for agents in games using social norms and emotions. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, IVA'11, pages 473–474, Berlin, Heidelberg. Springer-Verlag.

H. H. Vilhjalmsson, C. Merchant, and P. Samtani. 2007. Social puppets: Towards modular social animation for agents and avatars. In *Proceedings of the 2Nd International Conference on Online Communities and Social Computing*, OCSC'07, pages 192–201, Berlin, Heidelberg. Springer-Verlag.

H. H. Vilhjálmsson. 2004. Animating conversation in online games. In M. Rauterberg, editor, *Entertainment Computing ICEC 2004*, volume 3166 of *Lecture Notes in Computer Science*, pages 139–150. Springer Berlin Heidelberg.

H. H. Vilhjálmsson. 2014. Automation of avatar behavior. In J. Tanenbaum, M. S. el Nasr, and N. M., editors, *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters*, pages 255–266. ETC Press.

Biographical Sketch

Hannes Högni Vilhjálmsson is an Associate Professor of Computer Science at Reykjavik University where he directs the Center for Analysis and Design of Intelligent Agents (CADIA) and leads the Socially Expressive Computing (SECOM) group (`http:\\secom.ru.is`). Prior to RU, Vilhjálmsson was the technical director on the Tactical Language and Culture Training project at USC/ISI and one of the main architects of the early REA ECA and the BEAT nonverbal behavior toolkit at the MIT Media Lab. He can be reached at: `hannes@ru.is`.

Elías Ingi Björgvinsson is enrolled in the Language Technology Masters Program, an interdisciplinary program across University of Iceland and Reykjavík University. He has a BA in Theoretical Linguistics (Icelandic and German). He is a member of the Socialy Expressive Computing (SECOM) group where he works on spoken interaction in Icelandic. He can be reached at: `eliasb13@ru.is`

Hafdís Erla Helgadóttir is currently pursuing her BSc in Computer Science at Reykjavík University, where she just finished her first year. As a member of SECOM, she has been working on a research project over the summer concerning intelligent agents displaying dominant and submissive nonverbal behaviour and the effects of dominance on turn-taking. AI is of particular interest to her, inspired by her fascination with games and their making. She can be contacted at: hafdis13@ru.is

Stefán Ólafsson holds BAs in English and Chinese from the University of Iceland. He is currently a masters student in the Language Technologies program at the University of Iceland and Reykjavík University. He is a member of the Socially Expressive Computing group at CADIA where he focuses on dialog systems and cognitive architecture. His email is stefanola13@ru.is