Report on Representations for Multimodal Generation Workshop

Kristinn R. Thórisson, Hannes Vilhjalmsson, Stefan Kopp, and Catherine Pelachaud

■ The Representations for Multimodal Generation Workshop was held on April 23–25, 2005, at Reykjavík University, Reykjavík, Iceland. The overall goal of this workshop is to further the state of research on multimodal generation by enabling (and getting) people in the field to work together on building systems capable of real-time face-to-face dialog with people. This report summarizes the activities and progress of that meeting.

Human communicative behaviors span a broad set of skills, from natural language generation and production to eye gaze control and facial expression. People produce multimodal behavior with ease in real time in a broad range of circumstances. In April 2005 a group of practitioners in the area of multimodal communication and computer graphics came together for three days at Reykjavík University to further the integration and development of multimodal generation skills for artificial humans.

The group started by specifying the necessary planning stages and knowledge structures for multimodal behavior generation and control. While we felt it was important to form clear interfaces at separate levels of abstraction in order to modularize the system, we also did not want to force a particular microarchitecture on anyone. The solution was to identify abstraction levels at the macro level and treat the processing between them as black boxes or (more appropriately) as open research questions. This resulted in a focus on specifying the particular type and form of information that needs to be represented in the gaps between these levels.

We came up with what we call the *SAIBA framework* (situation, agent, intention, behavior, animation), according to which macroscale multimodal generation comprises three processing stages, or levels: (1) Intent planning, (2) behavior planning, and (3) behavior realization. The interface between intent planning (1) and behavior planning (2) describes communicative intent without any reference to physical behavior; the interface between behavior planning (2) and behavior realization (3) describes the physical behavior that carries out that communicative intent. The group analyzed how existing representation languages, APML, MURML, RRL, and older BML/FML, contribute to these interfaces and drafted new functional markup language (FML) and behavioral markup language (BML) specifications.

The group feels that a high degree of flexibility is achieved by separating the planning into these three levels in this particular way, and SAIBA is now being used as the foundation for further work by several of the attendees. Developing a common representation of the content to be exchanged by each planning level, without prescribing how the processing that produces these is done, opens the possibility of combining solutions from different researchers without too much code modification. In addition to the new FML and BML, which are still in development, the approach assumes additional development of a low-level representation of animation control below stage 3 and a high-level specification of goals above stage 1.

Unlike many such meetings, we explicitly structured the workshop to produce tangible results. The approach was considered a success; a follow-up is planned for next year.

The authors would like to thank the other participants, Norman Badler, W. Lewis Johnson, Stacy Marsella, and Brigitte Krenn, for their significant contributions and for making the workshop a success.

Kris Thórisson is an associate professor at Reykjavík University and the codirector of the AI lab (CADIA). He has been developing interactive AI systems for more than 15 years, both in academe and industry. His research focus is on multimodal cognitive robotics and sentient systems. Thórisson is cofounder of MINDMAKERS.ORG, an open effort to increase collaboration among AI researchers and significantly speed up the rate of progress in the field.

Hannes Vilhjalmsson is a research scientist at the University of Southern California's Information Sciences Institute in the Center for Advanced Research in Technology for Education. He is currently developing serious games for language and culture training that support social interactions with game characters.

Stefan Kopp is an assistant professor at the Artificial Intelligence Group at Bielefeld University, Germany. His research is centered around cognitive systems for intelligent human-machine interaction, with a focus on conversational characters. He is currently a research fellow in embodied communication at the Center for Interdisciplinary Research in Bielefeld and a member of the Collaborative Research Center 360 "Situated Artificial Communicators."

Catherine Pelachaud is a professor at the University of Paris 8, IUT of Montreuil, where she leads the Laboratory in Informatics and Communication (LINC). Her research interest includes representation languages for agents, embodied conversational agents, nonverbal communication (face, gaze, and gesture), expressive gesture, and multimodal interfaces.

