# A Shared, Modular Architecture for Developing Virtual Humans

Arno Hartholt [1], David Traum[1], Stacy Marsella[2], Louis-Philippe Morency[1], Ari Shapiro[1], and Jonathan Gratch[1]

[1]USC Institute for Creative Technologies, Los Angeles, USA
[2]Northeastern University, Boston, USA

## 1 Main Research Themes

Realizing the full potential of intelligent virtual agents requires compelling characters that can engage users in meaningful and realistic social interactions, and an ability to develop these characters effectively and efficiently. Advances are needed in individual capabilities, but perhaps more importantly, fundamental questions remain as to how best to integrate these capabilities into a single framework that allows us to efficiently create characters that can engage users in meaningful and realistic social interactions. This integration requires in-depth, inter-disciplinary understanding few individuals, or even teams of individuals, possess.

Our research is focused on understanding the relationship between individual capabilities, how they strengthen each other within larger systems and which sets of minimum and desired permutations can be defined for different types of systems and domains. This research is often conducted within the context of a common, modular framework that contains a mix of research and commercial technologies to offer full coverage of subareas including speech recognition, audio-visual sensing, natural language processing, dialogue management, nonverbal behavior generation & realization, text-to-speech and rendering.

Many of our characters are so-called *question-answering agents*. They offer a user-driven, interview-type style of conversation where a user question is answered with a character answer. These systems explore how to best provide particular information within a given domain, often through a set of pre-defined responses from one or more agents. Examples are SGT Star [1], Boston Museum of Science Guides [20], Virtual Patients [10] and Gunslinger [7], covering military, education, medical and entertainment domains.

Another focus has been on modeling verbal and nonverbal back-channeling behavior through *virtual listeners*, in order to establish rapport and increase speaker fluency and engagement. The Rapport project is an example of this [5].

*Virtual interviewers* put the focus on gathering information from or assessing the user. Shifting some of the conversational burden to the agent requires increased dialogue management and natural language understanding capabilities. The SimCoach system [17], for example, is a web-based guide that helps navigate healthcare related resources, which uses a forward looking, reward seeking dialogue manager [15]. Combining advanced dialogue management with audio-visual sensing results in SimSensei, an agent who aids in recognition of psychological distress [4]. SimSensei enables an engaging face-to-face interaction where the character reacts to the perceived user state and intent through its own speech and gestures.

Many of these capabilities are available for the research community through the ICT Virtual Human Toolkit[1]. It provides a solid basis for the rapid development of new virtual humans, but also serves as an integrated research platform to enable context-sensitive research in any of the Virtual Human subfields, taking advantage of and examining the impact on other system modules.

## 2 Current Architectures and Standards

Many of the virtual humans developed at the University of Southern California Institute for Creative Technologies are instantiation of a more general Virtual Human Architecture, see Figure 1. It defines at an abstract level the capabilities of a virtual human and how these interact. Not every system will include all capabilities and some will implement them to a greater or lesser extent. The architecture allows for multiple implementations of a certain capability and simple substitution of one implementation for another during runtime, facilitating the exploration of alternative models for realizing individual capabilities. Thus the general architecture can be specialized in many different ways.

Capabilities are realized through specific modules, most of which communicate with each other through a custom messaging system called VHMsg, build on top

[1] https://vhtoolkit.ict.usc.edu

of ActiveMQ[2]. Messages are typically broadcasted, although there are intended consumers. Libraries have been built for several languages, including Java, C++, C#, Lisp and TCL, so that developers have a wide latitude in developing new modules that can communicate with the rest of the system. There is a standard set of message types used by existing modules [6], and it is very easy to create new message types. The basis for this architecture is SAIBA[3] which was extended to cover additional areas. It uses the Functional Markup Language (FML) [9], the Behavior Markup Language (BML) [11], and the Perception Markup Language (PML) [21].

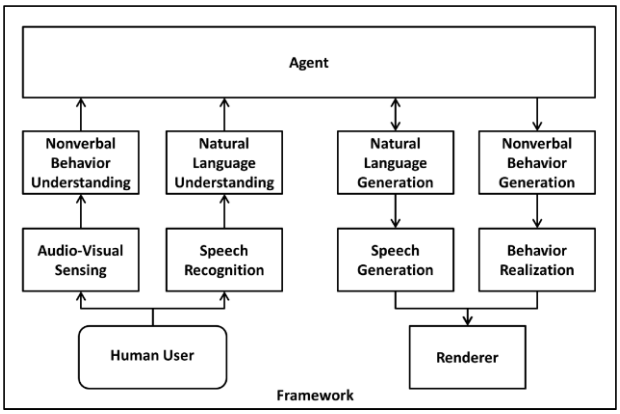


**Fig. 1.** The Virtual Human Architecture.

The relationship between capabilities and modules is not necessarily one to one. The Agent, for example, can range from rule-based systems [5], to a statistical text-classifier (the NPCEditor [13] which also serves as NLU and NLG), to cognitive architectures [23, 19], which can include sub-modules for task modeling, emotion modeling, and dialogue management. Other often used modules are MultiSense (audio-visual sensing) [21], FLoReS (dialogue manager) [15], Cerebella (nonverbal behavior generation) [14], and SmartBody (character animation) [22].

The strengths of the architecture lie in its modularity, extendibility and the wide range of integrated capabilities, offering a powerful basis for the rapid research and development of new implementations of both modules and systems. It has been validated by over two dozen systems, ranging from research prototypes to deployed applications. More information about the architecture, API, individual modules and capabilities, and created systems can be found in [6].

[2] http://activemq.apache.org/
[3] http://www.mindmakers.org/projects/saiba/wiki

## 3 Future of Architectures and Standards for IVAs

We suggest that future efforts focus on achieving broader interoperability between systems created by separate research groups. For instance, while several BML realizers exist [3, 8, 16, 22, 24], they are not directly interchangeable due to an often less than strict implementation of the standard as well as many custom extensions. In addition, the transport layer is undefined, preventing messages between in-house and external modules to be sent and received easily.

Related, more thought should be given to sharing capabilities, data and assets throughout the community, to more easily leverage each other's work. This requires that standards and procedures move beyond merely an architecture and interface, and include explicit notions of the community, their systems and data, and associated methodologies. A layered approach of Community, Architecture & Interface and Implementation levels may aid in this goal.

Reference architectures like SAIBA should include a broader range of capabilities, including audio-visual sensing and natural language processing. They should also address challenges like continuous communication and processing between components rather than the current often sequential ones.

Finally, relationships between capabilities should be made more explicit and should be extendable and scalable. At run-time, the system and its components should be able to detect available capabilities as well as their level of sophistication and adapt accordingly, analogous to discoverable web services. For instance, a speech recognition capability could offer either discrete or continuous recognition, with optional prosody analysis; the remainder of the system should be able to work with the minimum provided capability as well as take advantage of richer input when available.

## 4 Suggestions for discussion

As mentioned in section three, our interest is in creating an environment in which collaboration can happen more effectively and efficiently, both on a technical and organizational level. Relevant topics:

- Integrating architectures and standards from related fields (e.g. ROS, OpenInterface, OpenCog, EmotionML, the Incremental Unit-architecture,);
- Missing standards (e.g. Context Markup Language (CML), messaging, etc.);
- Data, assets, knowledge & technology sharing;
- Design and development best practices.

## References


1. Artstein R, Gandhe S, Leuski A, Traum DR. Field Testing of an interactive question-answering character. ELRA, LREC (2008)
2. Bickmore, T.W., Schulman D., Shaw, G., DTask & LiteBody: Open Source, Standards-based Tools for Building Web-deployed Embodied Conversational Agents, Intelligent Virtual Agents, PP. 425-431 (2009)
3. Julia Campbell, Matthew Hays, Mark Core, Mike Birch, Matthew Bosack, Richard E. Clark, Using Virtual Humans to Teach New Officers In Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) 2011
4. David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Alesia Egan, Ed Fast, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Skip Rizzo, and Louis-Philippe Morency. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), Paris, France, May 5-9, 2014.
5. Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, R. J. van der Werf, Louis-Philippe Morency, Virtual Rapport, IVA (2006)
6. Arno Hartholt, David Traum, Stacy C. Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, Jonathan Gratch, All Together Now: Introducing the Virtual Human Toolkit, at Intelligent Virtual Agents (IVA), 2013
7. Hartholt A., Gratch J., Weiss L., Leuski A., Morency L.P., Marsella S, At the Virtual Frontier: Introducing Gunslinger, a Multi-Character, Mixed-Reality, Story-Driven Experience IVA pp. 500-501 (2009)
8. Heloir, A. and Kipp, M.: A Realtime Engine for Interactive Embodied Agents, Intelligent Virtual Agents, pp. 393-404, 2009
9. Heylen, D.K.J., et al., The Next Step towards a Function Markup Language, at Intelligent Virtual Agents (IVA), 2008
10. Kenny, P.G., Parsons, T.D., Gratch, J. Rizzo, A., Evaluation of Justina: A Virtual Patient with PTSD, Lecture Notes in Computer Science, pp. 394-408 (2008)
11. Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson, Hannes Vilhjálmsson, Towards a Common Framework for Multimodal Generation: The Behavior Markup Language (LNAI), vol. 4133, pp. 205–217. (2006)
12. H. Chad Lane, Clara Cahill, Susan Foutz, Daniel Auerbach, Dan Noren, Catherine Lussenhop, William Swartout, The Effects of a Pedagogical Agent for Informal Science Education on Learner Behaviors and Self-efficacy, In Artificial Intelligence in Education, volume 7926, 2013
13. Leuski, A and Traum, D. NPCEditor: Creating virtual human dialogue using information retrieval techniques. AI Magazine, 32(2):42–56, (2011).
14. Stacy C. Marsella, Ari Shapiro, Andrew W. Feng, Yuyu Xu, Margaux Lhommet, Stefan Scherer, Towards Higher Quality Character Performance in Previz, Digital Production Symposium, Anaheim, CA, July, 2013
15. Morbini F., DeVault D., Sagae K., Gerten J., Nazarian A., Traum D., FLoReS: A Forward Looking, Reward Seeking, Dialogue Manager, Spoken Dialog Systems (2012)
16. I. Poggi, C. Pelachaud, F. de Rosis, V. Carofiglio, B. De Carolis, GRETA. A Believable Embodied Conversational Agent, Multimodal Intelligent Information Presentation, (2005)
17. Albert Rizzo, Eric Forbell, Belinda Lange, John Galen Buckwalter, Josh Williams, Kenji Sagae, David Traum, SimCoach: An Online Intelligent Virtual Agent System for Breaking Down Barriers to Care for Service Members and Veterans, Chapter in Healing War Trauma (2012)
18. Albert Rizzo, Bruce Sheffield John, Brad Newman, Josh Williams, Arno Hartholt, Clarke Lethin, John Galen Buckwalter, Virtual Reality as a Tool for Delivering PTSD Exposure Therapy and Stress Resilience Training, In Military Behavioral Health, volume 1, 2012
19. Paul S. Rosenbloom, The Sigma Cognitive Architecture and System, Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), 2013
20. Swartout W, Traum DR, Artstein R, Noren D, Debevec P, Bronnenkant K, et al. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. IVA pp. 286–300 (2010)

21. Stefan Scherer, Stacy C. Marsella, Giota Stratou, Yuyu Xu, Fabrizio Morbini, Alesia Egan, Albert Rizzo, Louis-Philippe Morency, Perception Markup Language: Towards a Standardized Representation of Perceived Nonverbal Behaviors, IVA pp. 455–463 (2012)
22. Shapiro A., Building a Character Animation System, 4th Annual Conference on Motion in Games 2011, Edinburgh, UK, November 2011
23. David Traum, William Swartout, Jonathan Gratch and Stacy Marsella, Virtual Humans for non-team interaction training, In International Conference on Autonomous Agents and Multiagent Systems (AAMAS) Workshop on Creating Bonds with Humanoids, 2005.
24. Van Welbergen, H., Reidsma, D., Ruttkay, Z.M., Zwiers, J.: Elckerlyc: A BML realizer for continuous, multimodal interaction with a virtual human. Multimodal UI (2010)

## Biographical Sketch

Arno Hartholt is a Computer Scientist at the University of Southern California (USC) Institute for Creative Technologies (ICT) where he leads the virtual human technology integration group as well as the Institute's central asset production & pipeline group. As such, he is responsible for much of the technology, art, processes and procedures related to virtual humans and associated systems. He has a leading role on a wide variety of research prototypes and applications in areas ranging from medical education and military training to intelligent tutoring and serious games. Hartholt studied computer science at the University of Twente in the Netherlands where he got his master's degree. He worked at several IT companies, from large multi-nationals to early start-ups, before accepting a position at USC ICT in 2005. As one of the main integration software engineers within the virtual humans project, Hartholt has developed a variety of technologies, with a focus on task modeling, natural language processing and knowledge representation. He can be reached at hartholt@ict.usc.edu.

David Traum is a principal scientist at ICT and a research faculty member at the Department of Computer Science at USC. At ICT, Traum leads the Natural Language Dialogue Group, which consists of seven Ph.D.s, four students, and four other researchers. The group engages in research in all aspects of natural language dialogue, including dialogue management, spoken and natural language understanding and generation and dialogue evaluation. The group collaborates with others at ICT and elsewhere on integrated virtual humans, and transitioning natural language dialogue capability for use in training and other interactive applications. Traum's research focuses on dialogue communication between human and artificial agents. He has engaged in theoretical, implementational and empirical approaches to the problem, studying human-human natural language and multi-modal dialogue, as well as building a number of dialogue systems to communicate with human users. He has pioneered several research thrusts in computational dialogue modeling, including computational models of grounding (how common ground is established through conversation), the information state approach to dialogue, multiparty dialogue, and non-cooperative dialogue. Traum is author of over 200 technical articles, is a founding editor of the Journal Dialogue and Discourse, has chaired and served on many conference program committees, and is currently the president emeritus of SIGDIAL, the international special interest group in discourse and dialogue. He earned his Ph.D. in computer science at University of Rochester in 1994.

Stacy Marsella has a Ph.D. from Rutgers University with a focus on AI and human problem solving. He is well known for his work in computational models of human cognition and emotion. He also has extensive experience in the design and construction of simulations of social interaction for a variety of research, education and analysis applications. This includes his work on virtual humans for immersive training environments such as ICT's MRE and SASO-ST systems and the DARPA-sponsored Tactical Language system. He leads several projects related to virtual humans, including SmartBody, a virtual human animation system, Cerebella, a nonverbal behavior generation system, and PsychSim, a model of social interaction based on theory-of-mind modeling as well as being co-developer of the EMA emotion model with Jon Gratch. He has also worked on psychotherapeutic applications of emotion models, including his work on Carmen's Bright Ideas, a system that teaches coping strategies to parents of cancer patients. Marsella plays a leadership role in organizing conferences on virtual humans, social intelligence and emotion modeling, has over 150 technical articles and is on the editorial boards of the Journal of Experimental And Theoretical Artificial Intelligence, IEEE Transac-

tions on Affective Computing and Journal of Intercultural Communication. He is member of the Association for the Advancement of Artificial Intelligence (AAAI), a fellow in the Society of Experimental Social Psychologists, and a member in the International Society for Research on Emotions.

Louis-Philippe Morency is a research assistant professor in the Department of Computer Science at the University of Southern California (USC) Viterbi School of Engineering and research scientist at the USC Institute for Creative Technologies, where he leads the Multimodal Communication and Machine Learning Laboratory (Multi-Comp Lab). He received his doctoral and master's degrees from MIT's Computer Science and Artificial Intelligence Laboratory. His research interests are in computational study of nonverbal social communication, a multi-disciplinary research topic that overlays the fields of multimodal interaction, computer vision, machine learning, social psychology and artificial intelligence. Morency was selected in 2008 by IEEE Intelligent Systems as one of the *Ten to Watch* for the future of AI research. He received six best paper awards in multiple ACM- and IEEE-sponsored conferences for his work on context-based gesture recognition, multimodal probabilistic fusion and computational modeling of human communication dynamics. His work has been featured in *The Economist*, *New Scientist* and *Fast Company* magazines.

Ari Shapiro has nearly two decades of professional experience in the computer field as an engineer, consultant, manager and scientist. He currently works as a research scientist at the USC Institute for Creative Technologies, where his focus is on synthesizing realistic animation for virtual characters. At ICT, he heads the team for the SmartBody application, which serves as an animation system for synchronizing speech, facial animation, body motion and gesturing for many of ICT's real time virtual human systems. For several years, he worked on character animation tools and algorithms in the research and development departments of visual effects and video games companies such as Industrial Light and Magic, LucasArts and Rhythm and Hues Studios. He has worked on many feature-length films, and holds film credits in The Incredible Hulk and Alvin and the Chipmunks 2. In addition, he holds video games credits in the Star Wars: The Force Unleashed series. Shapiro has published many academic articles in the field of computer graphics in animation for virtual characters, and is a five-time SIGGRAPH speaker. He completed his Ph.D. in computer science at UCLA in 2007 in the field of computer graphics with a dissertation on character animation using motion capture, physics and machine learning. He holds an M.S. in computer science from UCLA, and a B.A. in computer science from the University of California, Santa Cruz.

Jonathan Gratch's research focuses on virtual humans and computational models of emotion. He studies the relationship between cognition and emotion, the cognitive processes underlying emotional responses, and the influence of emotion on decision-making and physical behavior. He is the director for virtual humans research at the USC Institute for Creative Technologies, a research professor of computer science and psychology, and co-director of USC's Computational Emotion Group. He completed his Ph.D. in computer science at the University of Illinois in Urbana-Champaign in 1995. A recent emphasis of his work is on social emotions, emphasizing the role of contingent nonverbal behavior in the co-construction of emotional trajectories between interaction partners. His research has been supported by the National Science Foundation, DARPA, AFOSR and RDECOM. Along with ICT's Stacy Marsella, Gratch received the Association for Computing Machinery's Special Interest Group on Artificial Intelligence (ACM/SIGART) 2010 Autonomous Agents Research Award, an annual award for excellence for researchers influencing the field of autonomous agents. Gratch is the editor-in-chief of the journal IEEE Transactions on Affective Computing, a member of the editorial boards of the journals Emotion Review, and Journal of Autonomous Agents and Multiagent Systems. He is the former president of the HUMAINE Association for Research on Emotions and Human- Machine Interaction (now known as the Association for the Advancement of Affective Computing), and a frequent organizer of conferences and workshops on emotion and virtual humans. He belongs to the American Association for Artificial Intelligence (AAAI) and the International Society for Research on Emotion (ISRE). Gratch is the author of over 250 technical articles.

## Acknowledgments

We'd like to acknowledge the members of the USC ICT Virtual Humans group for their dedication and contributions. The effort described here has been sponsored by the U.S. Army. Any opinions, content or information presented does not necessarily reflect the position or the policy of the United States Government; no official endorsement should be inferred.